IMT School for Advanced Studies, Lucca

Lucca, Italy

Quantitative methods for computer aided decision support systems in Confocal Laser Endomicroscopy imaging of the gastrointestinal tract

PhD Program in Computer, Decision and System Science - Image Analysis

XXVIIIth Cycle

Davide Boschetto

2016

The dissertation of Davide Boschetto is approved.

Program Coordinator: Prof. Rocco De Nicola, IMT School for Advanced Studies Lucca

Advisor: Prof. Guido Caldarelli, IMT School for Advanced Studies Lucca

Co-advisor: Dr. Enrico Grisan, Università degli Studi di Padova

The dissertation of Davide Boschetto has been reviewed by:

Dr. Stefano Realdon, MD, Veneto Oncological Institute, Padova, Italy

Dr. Miguel Coimbra, University of Porto, Portugal

IMT School for Advanced Studies, Lucca

2016

Contents

List of Figures

List of Tables

Vita

January 22, 1987	Born in Padova, Italy
2009	B.Sc. in Biomedical Engineering Mark: 97/110 Università degli Studi di Padova
2012	M.Sc. in Bioengineering Mark: 100/110 Università degli Studi di Padova
July 2016	Big Data Scientist Previnet S.p.A., Preganziol (Treviso)

Publications

1. D. Boschetto, E. Grisan, "Quantitative methods for CLE images," *Working Paper*, 2016.
2. D. Boschetto, E. Grisan, "Automatic Segmentation of Villi in Confocal Endomicroscopy," *Working Paper*, 2016.
3. D. Boschetto, E. Grisan, "Superpixel-based classification of gastric chromoendoscopy images," *Accepted at SPIE Medical Imaging 2017*.
4. D. Boschetto, G. Gambaretto, E. Grisan, "Automatic Classification of Endoscopic Images for Premalignant Conditions of the Esophagus," at *SPIE Medical Imaging 2016*, vol. 9788, 2016.
5. D. Boschetto, G. Di Claudio, H. Mirzaei, R. Leong, E. Grisan, "Automatic Classification of Small Bowel Mucosa Alterations in Celiac Disease for Confocal Laser Endomicroscopy," at *SPIE Medical Imaging 2016*, vol. 9788, 2016.
6. D. Boschetto, H. Mirzaei, R. Leong, E. Grisan, "Superpixel-Based Automatic Segmentation of Villi in Confocal Endomicroscopy," at *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics, BHI 2016, Las Vegas, NV, USA, February 24-27, 2016*, pp. 168–171, 2016.
7. D. Boschetto, H. Mirzaei, R. Leong, E. Grisan, "Detection and Density Estimation of Goblet Cells in Confocal Endoscopy for the Evaluation of Celiac Disease" at *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp.6248–51, 2015.
8. D. Boschetto, H. Mirzaei, R. Leong, G. Tarroni, E. Grisan, "Semiautomatic Detection of Villi in Confocal Endoscopy for the Evaluation of Celiac Disease" at *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp.8143–6, 2015.

Other Publications

1. D. Boschetto, M. Castellaro, P. Di Prima, A. Bertoldo, E. Grisan, "Reconstruction of DSC-MRI data from sparse data exploiting temporal redundancy and contrast localization" in *Proceedings of the XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, vol. 41, pp. 225–228, 2014.

2. D. Boschetto, P. Di Prima, M. Castellaro, A. Bertoldo, E. Grisan, "Baseline constrained reconstruction of DSC-MRI tracer kinetics from sparse Fourier data" in *Proceedings of the 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pp. 321–324, 2014.
3. D. Boschetto, C. Rusu, R. Dharmakumar, S. Tsaftaris, "Temporal and Spatial Variation of Baseline Myocardial BOLD Signal Intensity in Cardiac Phase-Resolved BOLD MRI: A Potentially Revealing Insight into Dynamic Changes in Myocardial Oxygenation" in *Proceedings of the 23th Annual Meeting of ISMRM*, 2014.
4. C. Rusu, R. Morisi, D. Boschetto, R. Dharmakumar, S. Tsaftaris, "Synthetic Generation of Myocardial Blood-Oxygen-Level-Dependent MRI Time Series via Structural Sparse Decomposition Modeling" in *IEEE Trans Med Imaging*, vol. 33(7), pp. 1422–33, 2014.

Abstract

The mucosa of the gastrointestinal tract represents the main barrier between the inner body and the external world. A layer of cells runs from the esophagus to the rectum, playing a key role in preventing access to environmental hostile factors that could cause inflammation. Alterations in such mucosa are caused or can cause severe problems to patients, among others celiac disease, irritable bowel disease, Crohn's disease, ulcerative colitis and Barrett's esophagus. The gold standard for evaluating such diseases requires biopsies to be performed on the patient, often following the random four-quadrant protocol, other than positive serology. Quantitative methods for evaluating in-vivo these diseases, by exploiting distinctive image features that vary according to the grade of the disease, would improve the way clinical examinations are performed. This could in the long run lead to virtual biopsies with a single endoscopy examination. We propose a Computer Aided Decision Support System for endoscopic examinations performed using Confocal Laser Endomicroscopy for celiac disease and irritable bowel syndrome that, exploiting image features extracted in an automatic way, can assist the physician in its diagnosis and help him in selecting and identifying the areas that most require attention during an examination. Exploiting image features that are well-investigated in the literature, our tool outputs valuable information about the mucosa under examination with a friendly user interface. We hope with such solution to increase the attention towards the need of quantitative methods in this medical field.

Chapter 1

Introduction

1.1 Clinical importance of quantitative tools in endoscopy

Medical endoscopy is a minimally invasive and relatively painless procedure, that allows experts to inspect inner cavities of the human body. One important field in medical endoscopy is the inspection of the gastrointestinal (GI) tract, which is the main focus of this corpus of works. The gastrointestinal tract can be affected by severe forms of cancers with high mortality rates, and its mucosa can develop premalignant conditions such as Barrett's esophagus, adenomas, Crohn's disease and celiac disease. To diagnose with certainty such diseases, biopsy specimens analysis is required, on top of a positive serology of the patient under analysis. For flexible endoscopy, biopsies often do not correspond to the image under analysis, due to slight movement of the endoscope tip as a result of the preparation for taking a biopsy, especially in the case of magnified endoscopy. Digital imaging endoscopes, allowing the acquisition of images and videos, created a new field of computer aided decision support systems (CADSS) in medical endoscopy. Such systems are designed to detect and/or classify abnormalities in digital images, assisting the medical expert and actively improving the accuracy of medical diagnosis. Such systems usually follow the steps depicted in Fig. 1.

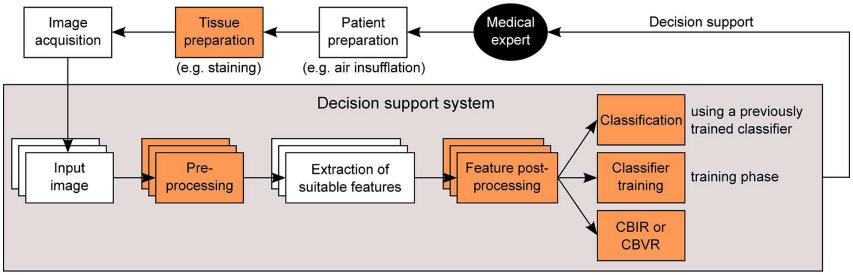

Figure 1: Pipeline of a general decision support system.

When images or videos are acquired, they are fed as input images to the system. A preprocessing step is usually performed to enhance the quality of the input images. After this step, suitable features must be selected, based on the purpose of the support system and on each image appearance. Based on these features, an automatic classification of the image is performed, using either a previously trained classifier or the features' values themselves, without resorting to classifiers. The classification results are given to the physician, providing him with additional information that could be useful to reach a better diagnosis. This additional information can be variegate, real time or on demand, based on the expert's will. Typically, the physician will have additional information about the content of the image under analysis, either with image segmentation-derived hints, or with feature-based classification methods on the whole image under analysis. Quantitative measures can be expressed and derived from the features extracted during this step, providing invaluable information to the expert. In fact, some premalignant diseases modify the appearance of the gastrointestinal mucosa (each disease on a specific sector of the GI tract). If such alterations could be measured and identified by using image features, diagnosis on certain premalignant diseases could be then done in vivo and in real time. It is already demonstrated [1; 2; 3] that confocal laser endomicroscopy (CLE), a recent technique that permits on site in vivo microscopic examination of the gastrointestinal mucosa, has a diagnostic accuracy comparable to

histology due to its high level of magnification. To successfully carry out in vivo histologies, quantitative methods are needed to establish with certainty if a patient shows problematic symptoms.

1.2 Problem statement

At the moment, no computer aided decision support system has been proposed for confocal laser endomicroscopy images. CLE is still a relatively novel technique, with quantitative methods exploiting its unique characteristics still missing from the literature. Although this technique is still under investigation and not widely adopted, it has all the features to produce a ground breaking impact on medical endoscopy, since virtual biopsy will be feasible using such technique. Computer-assisted diagnosis is not an easy goal in the very challenging setting that is the gastrointestinal tract. With flexible endoscopy, data-sets are constrained, images can show degradation and artifacts, and the appearance of the duodenal tissue is itself highly dependant on camera parameters. Angle and distance from the mucosa can cause blurred, underexposed or noisy areas of the image [4; 5]. Current gold standard might suffer from patchy distribution of intestinal mucosa areas among normal mucosa [6; 7]: if biopsies are taken only from the healthy mucosa due to alignment and sampling errors, the proper diagnosis will be missed. The air insufflation and suction and the instillation of water into the lumen cause bubbles in the walls, reflections of the light sensor installed in the endoscope, that both sensibly affect the feature extraction process and the performances of CAD systems, blur and noise especially [8]. Existing CADs for flexible endoscopy (not for CLE) are tailored toward specific applications, therefore are hardly adaptable for the tasks that we are interested in (diagnosing and staging CD and IBS). On top of this, during flexible endoscopy, the endoscope ofter travels through stomach and duodenum multiple times: these two different sectors of the GI tract have similar visual appearance. This can be a problem for a CAD system tailored for CD diagnosis with normal endoscopy: in fact, esophageal and gastric mucosa could be mistakenly labeled as CD-positive duodenal mucosa [5; 9] (see

Figure 2: Visual appearance of different GI regions with white light endoscopy: esophagus, stomach and CD duodenal mucosa.

Fig. 2), given that they are almost impossible to distinguish with a classification approach purely based on image features.

1.3 Thesis organization

In this work,quantitative methods exploiting the unique characteristics of images obtained through CLE have been devised and implemented: these methods will help physician and, in the long run, will make virtual biopsies possible.

In Chapter 2, the reader will be introduced to the necessary background information and perform a literature review of the state of the art methods in medical endoscopy. In detail, computer aided decision support system will be introduced and reviewed, and details about how they work and why they exist will be presented. Then, the reader will be introduced to the most critical clinical problems of the gastrointestinal tract that have been taken into account in this work: celiac disease and irritable bowel syndrome. Then, a comparison among different methods of imaging the GI tract will be performed, by comparing and analyzing the differences among white light endoscopy imaging, narrow band imaging, confocal laser endomicroscopy and wireless capsule endoscopy. Then, to conclude the background chapter, a quick review of existing state of the art CADSS in medical endoscopy will be presented: most of the existing methods are used for polyp detection and for image

enhancement.

In Chapter 3, the main scientific contributions will be presented. First, another brief introduction to the level of details and to the clinically relevant information that can be extracted from a CLE image will be given to the reader. Then, a fully automatic method for detecting and segmenting villi in CLE images will be proposed. This step is the basis from which all the quantitative analysis have been built upon. From this information, then, an algorithm to detect, segment and highlight goblet cells within villi is presented and detailed: goblet cells' density is, in fact, among the clinically relevant features that physicians look into for staging celiac disease. Then, an algorithm devised to detect crypts in CLE images is presented, since crypt hyperplasia is another features that correlates with severe cases of celiac disease, along with villous atrophy. Automatic ways of detecting fluorescein leakage and cell dropout in CLE images from the mucosa folds are then presented and detailed. These features are clinically established features that physicians evaluate when diagnosing patients with irritable bowel syndrome. Finally, all these algorithms are put together to give quantitative metrics to the physician (goblet cells' density, crypt density, fluorescein leakage and/or cell dropout) to support him in his diagnosis. For this, an easy-to-use software with a graphical user interface (GUI) has been developed in MATLAB, to give the medical expert access to the proposed tools at a high level. This software will be shown in Chapter 4.

Chapter 2

Background and Literature Review

2.1 The gastrointestinal tract

The large, muscular tube that extends from the mouth to the anus is called the gastrointestinal tract (GI tract). The movement of muscles (peristalsis), along with the release of hormones and enzymes, allows for the digestion of food. Its main sectors are the esophagus, the stomach, the small intestine and the large intestine. Each of these serve a different purpose, and therefore present different anatomy and structures in their mucosa.

2.1.1 Esophagus

Usually 18-25 centimeters long, the esophagus (or oesophagus) is an organ consisting of a fibromuscular tube connecting the mouth to the stomach, from which the food passes aided by peristaltic contractions. The wall of the esophagus (from the lumen outwards) consists of mucosa, connective tissue (submucosa), muscle fibers between layers of fibrous tissue and connective tissue. The mucosa is formed by three layers of squamous cells (stratified squamous epithelium), in contrast to

the single-layer columnar epithelium that is found in the stomach. The transition among these two types of epithelium is known as the gastroesophageal junction (GEJ), and is a visible zig-zag line (called z-line). Two muscular sphincters prevent reflux. The typical pink color of the esophageal mucosa contrasts to a deeper red of the gastric mucosa. Usually, the z-line corresponds with the upper limit of the gastric folds of the cardia, but the anatomy of the mucosa can be distorted (as in Barrett's Esophagus). The mucosa has no keratin and has a smooth lamina propria. Its epithelium serves as protection against the abrasive effect of food, while glands secrete mucus (esophageal glands in the submucosa and esophageal cardiac glands in the lamina propria).

2.1.2 Stomach

The stomach's main function is to secrete protein-digesting enzymes (proteases) and gastric acid to aid in food digestion following chewing. The chyme (partially digested food) is then transferred to the small intestine. Two sphincters keep the content of the stomach contained. Its near-empty volume in adult humans is of about 75 milliliters. Being a distensible organ, it normally expands to a volume in which it can hold one litre of food. Anatomically, it is divided in four sections:

- the first (cardia) is the region following the z-line;
- the fundus, formed by the upper curvature of the organ;
- the body, composed by the main central region;
- the pylorus, the lower section of the organ facilitating the emptying process towards the small intestine and the duodenum.

Gastric epithelium consists of simple columnar tissue, organized into gastric pits and glands to deal with secretion. The submucosa consists of fibrous connective tissue.

2.1.3 Small bowel

Divided in duodenum, jejunum and ileum, the small bowel (also, small intestine) is where most of the digestion and absorption of food take place. The surface area of the small bowel mucosa averages $30\ m^2$, given its length on average of about 7 m and its diameter of $2.5 - 3$ centimeters. The duodenum is continuous with the stomach and C-shaped. It mixes the gastric chyme and the digestive juices from the pancreas and the liver (digestive enzymes and bile) which break proteins and emulsify fats. A mucus-rich alkaline secretion containing bicarbonate is produced and mixed with chyme in the duodenum, neutralizing the gastric acids. The jejunum is about 2.5 m long. It contains plicae circulares and villi, that drastically increase its surface area. Here all the products of the digestion process (sugars, amino acids and fatty acids) are absorbed into the bloodstream. The ileum is about 3 meters long and contains villi as does the jejunum. It is specialized on absorption of vitamin B12 and bile acids, as well as other remaining nutrients. At the ileocecal junction, it joins the large intestine. Intestinal epithelium contains goblet cells (in charge of secreting mucin). Thanks to the villi (and microvilli, finger-like projections of epithelial cells) nutrients are transported from the intestinal lumen into capillaries and lacteals, absorbed, and transported via those blood vessels to different organs of the body.

2.1.4 Large bowel

Large bowel (also called colon or large intestine) is the last part of the digestive system, whose main purpose is water and salt absorption. Overall, it is about 1.5 m long, about one fifth of the whole gastrointestinal tract. It is composed of, starting from the small intestine, the cecum, the colon (ascending, transverse, descending and sigmoid), the rectum and the anal canal. In the ascending colon, the unwanted waste material is moved upwards through peristalsis. Here, the process of absorption of the water starts. The colon shows simple columnar epithelium with invaginations, called intestinal glands or colonic crypts. In the 16 hours that the large intestine needs to finish the digestion process, water and

salt are removed from the food before storing the indigestible matter to the rectum. Colon absorbs vitamins that are created from the colonic bacteria, such as vitamin K, vitamin B12, riboflavin and thiamine. Feces are compacted and stored in the sigmoid colon, which has muscular walls, that can contract to increase colon's pressure, pushing the stool towards the rectum. No digestive enzymes are produced in the large intestine: chemical digestion is in fact completed in the small intestine, before the chyme reaches the large intestine. The pH is slightly acidic to neutral (5.5-7).

2.2 Celiac Disease and CLE

Celiac disease (CD) is one of the most frequent immune-mediated enteropathies that affects genetically susceptible persons triggered by exposure to gluten (a protein found in wheat, rye, barley, vitamin and nutrient supplements and certain medications) and similar proteins [10; 11; 12]. CD has a prevalence ranging from 1:80 to 1:300 in Europe and in North America [13; 14; 15; 16], even though this disease is considered a hidden epidemic. It is estimated in fact that most of the celiac patients (seven out of eight patients suffering CD) will remain undiagnosed during their life [17], since most CD cases only have minor gastrointestinal symptoms [18]. Celiac disease is more common in Caucasians and females, and among people with Down syndrome and Turner syndrome. CD causes a wide spectrum of clinical manifestations: gastrointestinal symptoms such as abdominal pain, chronic diarrhea, bloating, nausea, constipation and vomiting, and other general symptoms such as weight loss, fatigue, failure to thrive, slowed growth, short stature, dermatitis [19], alopecia areata, cerebellar ataxia [20], iron deficiency [21], anemia, seizures, and premature osteopenia, among others. Signs and symptoms vary this much since they depend on numerous factors. These include the age a person started eating gluten and in what amount, age of the person and small intestine damage degree. Long term complications include, as said, malnutrition along with liver diseases, intestinal cancer and lymphoma. Exposure to gluten causes variable damage to the small

bowel mucosa: mild damages include cases with increased number of intraepithelial lymphocytes and the presence of Crypt Hyperplasia (CH), while severe forms of the lesions involve various degrees of endoscopically relevant lesions such as villous atrophy (VA) [22], in which finger-like structures are projected outside the mucosa. Villi normally absorb nutrients from foods and pass the nutrients through the small intestine into the bloodstream. Without healthy villi, people risk malnourishment, no matter how much food they eat. If a patient is inappropriately treated for CD, the inflammation caused by the deregulated immune response can therefore disrupt the intestinal mucosa. This disruption process prevents the ability of the mucosa to absorb nutrients. Sometimes, some of the symptoms are similar to those of other diseases, therefore CD might mistakenly be diagnoses as irritable bowel syndrome, lactose intolerance, diverticulitis or chronic fatigue syndrome. The patients need at first to adopt a strict gluten free diet (GFD) as the first choice treatment modality, to suppress inflammations allowing therefore a healing process of the mucosa. Patients need to maintain a strict GFD diet for the remaining of their life, if they want to avoid acute or chronic complications of CD [23]. CD-induced alterations of the mucosa are not easily recognized during normal flexible endoscopy. Thus, in everyday practice, the identification of CD is made on the basis of a positive diagnostic intestinal biopsy and of the concomitant presence of a positive celiac serology [24]. The gold standard in the diagnosis of CD is the demonstration of VA in duodenal biopsies [25]. VA is investigated extensively in the medical community [26; 27; 28], and image processing methods as well as quantitative computational methods are highly needed, required and recommended from the community for the characterization of the small intestinal mucosa in suspected and known CD patients [29]. A blood test can be performed on a patient to test for the presence of antibodies typical of CD, but if a patient stops eating foods with gluten before being tested, the results may be negative for CD even if the disease is present. Therefore, the gold standard requires intestinal biopsy (it is performed after a positive blood test, too, to confirm the preliminary diagnosis). Mucosal alterations caused by CD are classified into

different stages of severity [30; 31], even though the histological staging of biopsies is subject to significant degrees of intra- and inter-observer variability [32; 33; 34; 35]. For this reason, observer independent diagnostic methods (such as CADSS) are needed [9]. On top of this, the whole process is time consuming, cost intensive and invasive. Diagnostic methods are observer-dependent, and require significant knowledge, expertise and time [36; 37].

2.3 Intestinal permeability-related diseases

The mucosa of the gastrointestinal tract represents the main barrier between the inner body and the external world. A layer of cells runs from the esophagus to the rectum, playing a key role in preventing access to environmental hostile factors that could cause inflammation. In particular, the intestinal epithelium is the largest mucosal surface, regulating the transit of macromolecules [38; 39]. This barrier is formed by a double layer of lipid cells, offering strong resistance to water soluble constituents. The junction between epithelial cells is a region in which inter-cellular junctions (tight junctions) are formed, to regulate the constituents' flow. This junction's permeability is dynamic, varying according to dietary state, humoral or neural signals and inflammatory mediators, among others. If pathological conditions ensue, the permeability is increased and a loss of epithelial integrity is suffered. Impaired epithelial barrier function is present in both typical inflammatory bowel diseases (Crohn's disease and ulcerative colitis [40; 41]) and in irritable bowel disease (IBS), a somehow still mysterious functional disorder.

2.3.1 Inflammatory bowel diseases

The most common inflammatory bowel diseases are ulcerative colitis and Crohn's disease, and both involve chronic inflammation of part of the digestive tract. Both diseases involve severe diarrhea, pain, fatigue and weight loss, and can be debilitating and lead to life-threatening complications. Crohn's disease usually affects the small intestine and the

initial part of the large intestine, though it can affect any part of the GI tract (from the mouth to the anus). Its causes are still unknown, but it is believed that autoimmune reaction, along with environmental (antibiotics, oral contraceptives [42], high-fat diet [43]) and genetic factors may play a role in causing it. Briefly, bacteria or virus might mistakenly trigger the immune system to attack the inner lining of the intestines, which causes inflammation, leading to symptoms. The disease is more likely to develop in young smokers (especially 20-29 year-olds [44]) with a family member with IBD. It is diagnosed with blood and stool tests, and with an intestinal endoscopy with biopsy samples analysis. This disease can be treated with medications, bowel rest or surgery, but each person will experience Crohn's disease differently. Usually, immunomodulators and/or corticosteroids are used to alleviate symptoms, bowel rest (with intravenous nutrition) is suggested for the intestines to heal, but about one person every five will need surgery [45] to treat complications and improve symptoms, while not curing Crohn's disease. Small bowel resection, subtotal colectomy or proctocolectomy and ileostomy are the general available choices for the surgery path. Eating, diet and nutrition don't cause Crohn's disease, but a good nutrition plan is important in its management, though. Generally, carbonated drinks need to be avoided, along with high fiber foods. Liquid intake is very important, and eating smaller meals more often is suggested to ease the symptoms. Typical complication of the disease, if left untreated, are bowel obstruction, fistulas (abnormal passages between two organs or an organ or the outside world), ulcers, malnutrition or general inflammation in joints, eyes or skin. A correlation between Crohn's disease and colon cancer has been proved, too [46; 47]. Ulcerative colitis, instead, causes sores called ulcers on the inner lining of the large intestine. Usually, this disease become worse over time, with symptoms ranging from mild to severe, although remissions are typical, lasting for weeks or years. The goal of healthcare is to keep patients in remission for as much as possible. Again, specific causes for ulcerative colitis are unknown, but it is supposed that an overactive intestinal immune system (along, again, with genetic and environmental factors) is the main factor inducing UC.

This disease is more likely to develop in people of Jewish descent, with family history of IBDs, with an age either of more than 60 years old or between 15 and 30 years old [48]. The most common symptoms are diarrhea with blood or pus, and abdominal discomfort. Of all the affected patients, about 10% can have severe symptoms [48] and complications (rectal bleeding, dehydration and malabsorption). As is the case with Crohn's disease, endoscopy is the main method for diagnosing ulcerative colitis, and medications can only reduce its symptoms. Surgeries are often performed only if the patient shows colon cancer, precancerous cells in the colon (dysplasia), life threatening complications (megacolon or bleeding), no symptom improvement despite treatment, dependency on steroids or side effects from the medications. Sometimes, a removal of the entire colon is required, with a full recovery that takes 4 to 6 weeks with using an ostomy pouch for the rest of the patient's life to collect intestinal contents. The patient should follow the same dietary suggestions that are given for Crohn's disease.

2.3.2 Irritable bowel syndrome

Irritable bowel syndrome is not an inflammatory disease, but it is a functional disorder. Its typical symptoms are abdominal pain, constipation (mostly in women) and diarrhea (mostly in men). Some people lament feeling very full, flatulence of mucous discharge, along with cramps. This disease is known with other names too, such as irritable colon, mucous colitis, spastic colon, spastic colitis and nervous stomach. IBS in itself is not dangerous, but its symptoms are painful and bothersome, and can affect people's everyday activities, causing distress and actively affecting the standard of living of the patients suffering from it. Most people have a mild form of IBS, that can go untreated with acceptable symptoms. Mostly, IBS affects people between 35 and 50 years old, with about 14% of women and 9% of men affected. It is suspected that IBS is caused by oversensitive nerves in the intestine, intestinal muscle disorders, bacterial flora imbalance and/or inflammations of the intestinal wall, possibly along with genetic predisposition. All of these causes, since IBS

mostly affects the large intestine, alter the speed in which the food mass transverses the intestine: if it goes through too quickly, the food mass does not lose enough water. On the other hand, with a slow passage, people might suffer from constipation and abdominal cramps. As of today, there is no cure for IBS. Though, by experimenting and paying attention to lifestyle and food habits, most people find out what makes them feel better or worse. The criteria that should lead to a positive diagnosis are debated, but generally a person is considered to have IBS if:

- symptoms as stomach ache or flatulence have lasted for more than three months, caused by the bowel with changes in the stool;
- a person is significantly impacted on its quality of life due to the symptoms;
- symptoms are certainly not caused by another disease.

Symptoms, though, can be caused by other diseases such as lactose intolerance, gluten intolerance and celiac disease, and a patient could suffer from more than one diseases at the same time.

2.4 Imaging the Gastrointestinal tract

It is clear that multiple different modalities exist to image the gastrointestinal mucosa. Each image modality has its own strength and weak spots: in this section each of the most used modalities (MR and CT imaging, white light endoscopy, narrow band imaging, chromoendoscopy, autofluorescence imaging, wireless capsule endoscopy and confocal laser endomicroscopy) will be analyzed.

2.4.1 MR and CT imaging

MRI can be used in the gastrointestinal tract for imaging and for measuring interesting markers, such as intestinal motility indexes, small bowel water content and whole gut transit time [49]. As of today, though, these MRI bio-markers are not included in the clinical routine, because of the

expensiveness of the technique, and because they are still not proved reliable by enough trials in the clinical practice. Regarding imaging, bowel examinations can be performed by combining the use of T2-weighted single-shot and breath hold T1-weighter gradient echo, with or without fat suppression, and gadolinium-enhanced 3D gradient echo. True-FISP sequences can detect bowel wall pathology and overall bowel anatomy, with robustness to motion, if combined with water-based intraluminal distending and contrast agents. This technique, though, suffers from artifacts (due to the extreme sensitivity to field inhomogeneities), and can not image well with the presence of retained bowel gas, in fact impairing bowel wall assessment [50]. CT, instead, has been used for virtual endoscopy for the large bowel, but since it is less cost-effective than conventional endoscopy it has not been adopted widely in the medical practice. On top of this, the irradiation exposure is relatively high during these examinations [51].

2.4.2 Common flexible endoscopy modalities

Historically, the first endoscopic image modalities has been white light endoscopy. In the 1960s, flexible endoscopy was introduced using fiber optics and a lens to visualize the inner cavities of the human body. Modern endoscopies are compact devices that are able to record and store digital high resolution images thanks to a light source and a CCD/CMOS sensor. Using an accessory channel allows physician to use medical instrument to sample tissue specimens (for biopsies), to clean and prepare areas under analysis and perform polypectomies without invasive surgery. In Fig. 3 the typical setting of a normal WLE exam is shown: the mucosa is imaged by the endoscope, with an optional zoom (zoom-endoscopy) available for certain endoscopes.

In common flexible endoscopy modalities, the physician has the control on the endoscope, and moves it according to the feedback from the video feed. Typical field of view is of about 170 degrees, and no post-processing is performed on the images acquired by the sensor. A typical image acquired via WLE is shown in Fig. 4. Chromoendoscopy requires a

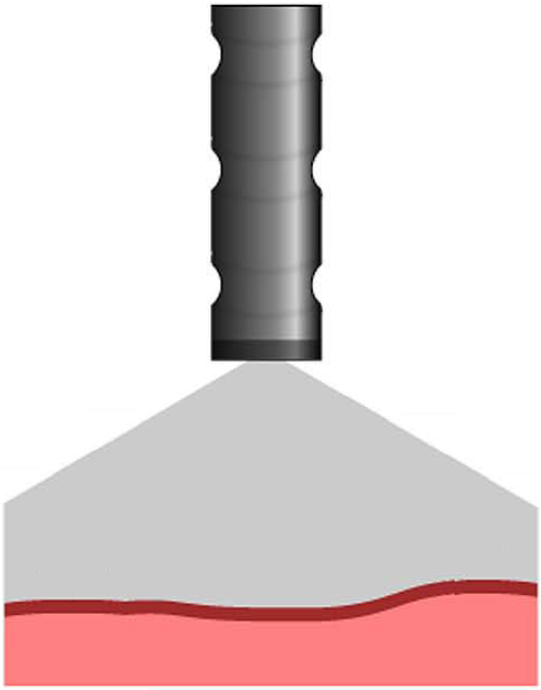

Figure 3: Typical setting of a WLE exam. The endoscope images the mucosa within a certain field of view, at a certain distance from the mucosa.

Figure 4: Typical image resulting from WLE, from a patient suffering from Barrett's Esophagus, a premalignant disease of the lower esophagus.

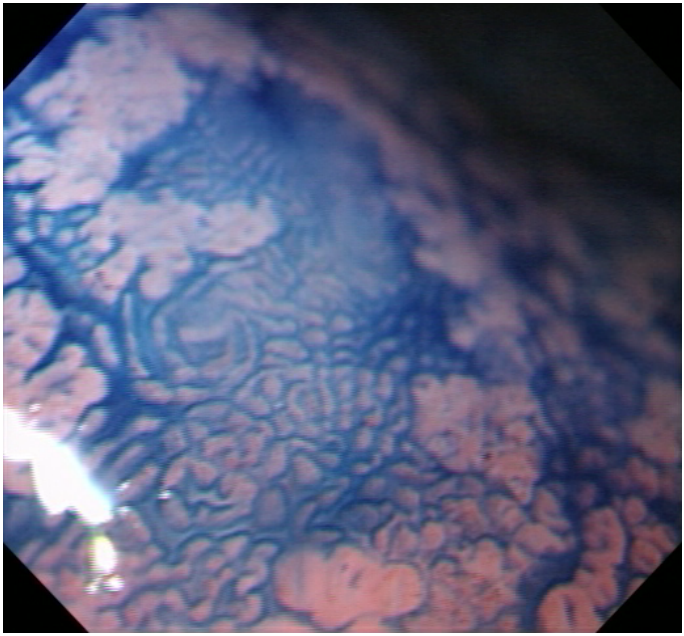

Figure 5: Typical image resulting from chromoendoscopy, from a patient suffering from suspect gastric cancer, stained with methylene blue.

normal WLE endoscope, but uses dyes at the time of visualization. Dyes enhance tissue characterization, uncovering texture pattern details that would not be otherwise clear without their use. Different stains can be used, according to the mucosa under investigation. Absorptive stains are useful for characterizing squamous epithelium, therefore highlighting dysplasias or metaplasia. Contrast stains are not absorbed and provide a contrast enhancement, while reactive stains test some functionalities of the GI tract, due to chemical reactions at certain pHs. This image modality, though, is time consuming and not easy to use as WLE. A typical image acquired via CHR is shown in Fig. 5.

To compensate for this, virtual chromoendoscopy saw an increase in its adoption in the latest years. Fuji Intelligent Chromoendoscopy (FICE) and, more generally, narrow band imaging (NBI) exploit physical prop-

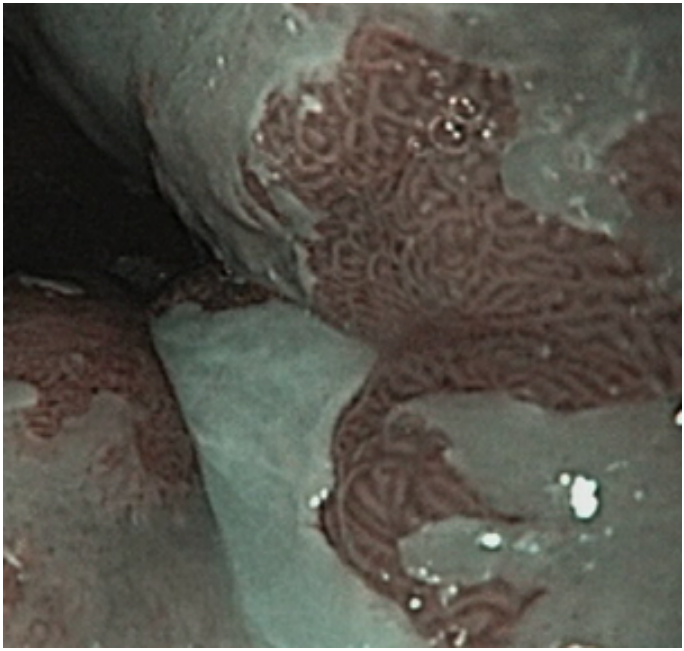

Figure 6: Typical image resulting from NBI, from a patient suffering from Barrett's Esophagus.

erties of the mucosa to enhance differences among different tissues. Instead of using white light, NBI uses light of specific blue and green wavelengths. This special image filter is activated by a switch in the endoscope, leading to the usage of light with wavelength between 440 and 460 nm and one of wavelength between 540 and 560 nm. The chosen wavelengths are the ones in which hemoglobin has its peak light absorption, therefore rendering blood vessels very dark, improving at the same time identification of different tissues. A typical image acquired via NBI is shown in Fig. 6.

A relatively novel technique is autofluorescence imaging (AFI), that has gained traction in the last few years in the medical endoscopy field. This image modality is based on the detection of natural tissue fluorescence emitted by endogenous molecules (fluorophores) such as col-

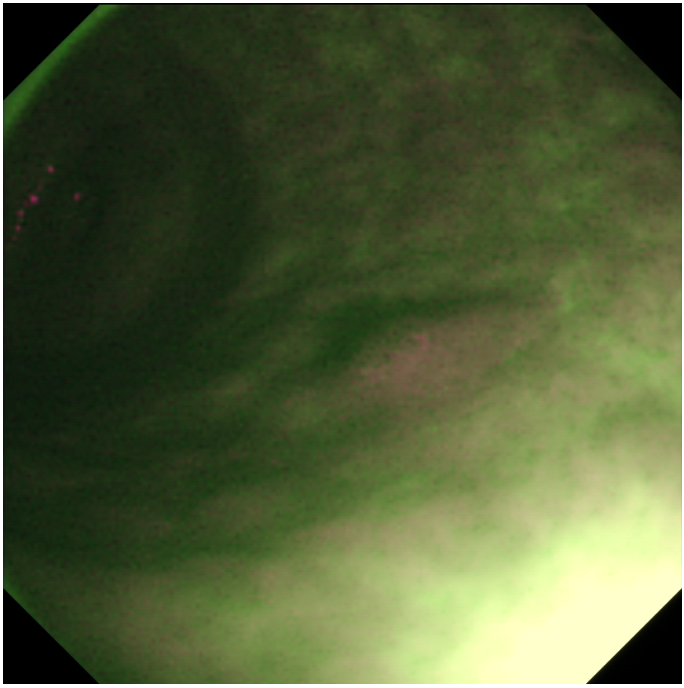

Figure 7: Typical image resulting from AFI, from a patient exhibiting a polyp in the mucosa (not showing the green autofluorescence in this image).

lagen [52]. The fluorescence emission differs among various tissue types due to corresponding differences in fluorophore concentration, metabolic state and/or spatial distribution. These differences can be captured in real time and used for lesion detection or characterization [53].

AFI is used as a part of a trimodal imaging video endoscopes using the RGB sequential illumination platform. Two CCD sensors are located at the tip of these endoscopes: one is dedicated to high definition WLE and NBI, while the other is assigned to AFI. A switch on the endoscope gives the medical expert easy possibility of switching among the three different image modalities (Olympus EVIS LUCERA SPECTRUM, Tokyo, Japan). AFI implements a special rotating color filter wheel in

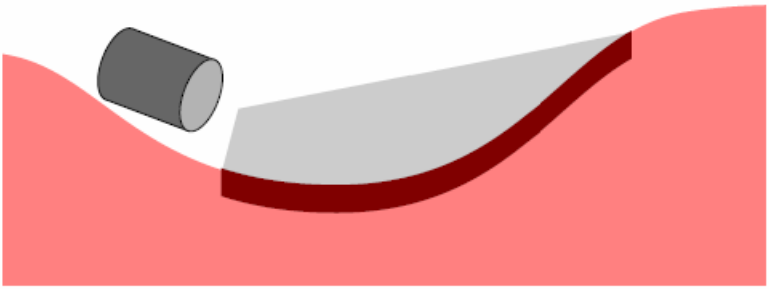

Figure 8: Typical setting of a WCE exam. The capsule travels through the GI tract, without any motion control.

front of the light source (xenon), that sequentially generates blue and green light (390-470 nm and 540-560 nm, respectively) for tissue illumination. Then, a filter in front of the AFI CCD blocks blue light excitation but enables autofluorescence (500-630 nm) and green light to filter through. The processor encodes a real time pseudo-color image in which normal mucosa appears green, while dysplastic tissue usually appears dark purple (as shown as example in Fig. 7). AFI is a valuable tool when used in a multi-modal imaging scheme, but lacks specificity to be used in an useful way as a stand-alone diagnostic tool [52].

2.4.3 Wireless capsule endoscopy

Wireless video capsule endoscopy (WCE or VCE) is an imaging technique that has appeared in 2001 in the medical endoscopy field, used mainly for investigating small bowel pathology, and to determine the cause of obscure gastrointestinal bleeding, Crohn's disease, polyposis and complications of CD [54].

It consists of an ingestible camera in a pill which passes through the patient's GI tract, taking 50.000-60.000 digital images in 8-12 hours. The capsule is about 26×11 mm in size, consisting of an optical dome, a lens, some emitting diodes, a transmitter and an antenna. The antenna transmits the digital pictures to a data recorder, worn externally by the patient,

where images are stored. After the study, the images are viewed and interpreted by a specially trained gastroenterologist. The main sector of the GI tract that is investigated with this technique is the small bowel, although esophagus and colon can be investigated, too (usually with some variations of the capsule, with two symmetric domes and cameras for the esophageal case, for example). The capsule is physiologically propelled through the entire GI tract, with bowel movements (an image representation of its imaging scheme is shown in Fig. 8). Being absolutely non invasive is the main advantage of this technique: it is considered a safe procedure since the capsule is usually expelled by the host within 24-48 hours from the exam. In case of retention not due to small bowel obstructions (if the patient experiences obstructions, WCE should obviously not be performed in the first place), a case report confirmed that no harm or dangerous symptoms were suffered by the patient [55]. No sedation is needed for the procedure, which is completely painless. Other than people with small bowel obstructions, patients with swallowing disorders or carrying pacemakers or other implanted cardiac devises (other than pregnant women) are also advised to have a careful specialist evaluation before undergoing capsule endoscopy. The patient undergoing capsule endoscopy can proceed with his normal life activities. This image modality is separated from flexible endoscopy (and CLE) since the physician has no direct control over the imaging device. The process is very time consuming, given the quantity of images that need to be accurately reviewed by an expert, and CADs have been proposed in the literature for this endoscopic modality [56].

2.4.4 Confocal laser endomicroscopy

CLE is an image modality developed to obtain in vivo very high magnification and resolution images of the mucosal layer of the GI tract. While conventional endoscopy needs biopsy for histological analysis, confocal endomicroscopy allows for in vivo histological examinations of tissue (a visualization scheme of how it works is shown in Fig. 9).

The advent of this new endoscopic optical technique is likely to "change

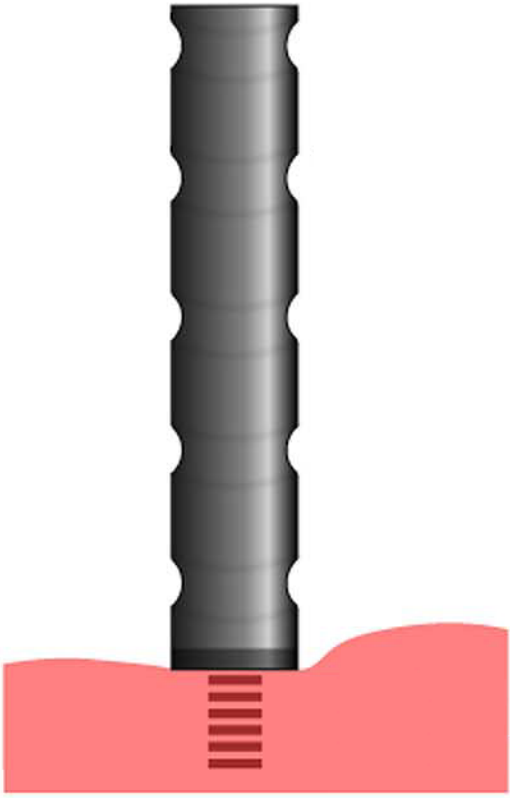

Figure 9: Typical setting of a CLE exam. The confocal tip fires a low-power laser, resulting in a high resolution picture of the chosen focal plane.

Table 1: Specifications of eCLE and pCLE in the upper and lower GI

	eCLE	pCLE
Distal diameter, mm	6.3-12.8	2.6 or 2.7
Field of view, μm	475	240 or 600
Imaging depth, μm	0-250	70-130
Lateral resolution, μm	0.7	1-3.5
Axial resolution, μm	7	15
Image rate, images/s	0.8-1.6	9-12

pathologists' role in diagnosis" [57].

CLE was developed in 2004 by Pentax, and cleared for use by FDA in the same year. Two different systems for CLE are available: endoscope-based CLE (eCLE, by Pentax) uses a fiber optic cable to convey blue laser light to a miniaturized confocal microscope integrated into the 12-mm-diameter tip of an endoscope [58], and probe-based CLE (pCLE, by Mauna Kea Technologies [59]), which uses a fiber-optic probe bundle to convey light from a confocal microscope situated outside the patient to a port in a standard endoscope [60]. For a comparison among the two systems, refer to Tab. 1. Both platforms exclude light from planes above and below the plane of interest using the pinhole aperture technique. Briefly, as shown in Fig. 10, the laser light is focused on an area of interest and back-scattered light is then refocused onto the detection system by the lens. The back-scattered light passes through a pinhole aperture which increases the resolution of the image.

This process needs a dye to provide contrast for accurate visualization, since CLE is based on tissue exposure and fluorescence and that in most cases tissue autofluorescence does not provide sufficient contrast [61]. Therefore, exogenous contrast agents need to be applied, either intravenously or topically. The most widely used CA is five milliliters of intravenous 10% fluorescein (Fluorescite; Alcon, Switzerland). Fluorescein IV is approved by FDA for retinal angiography, but it has found widespread use off-label in CLE. Fluorescein is chemically closely related to eosin: they produce essentially identical staining of cytoplasm and extracellular matrix. Fluorescein also highlights the structure and phys-

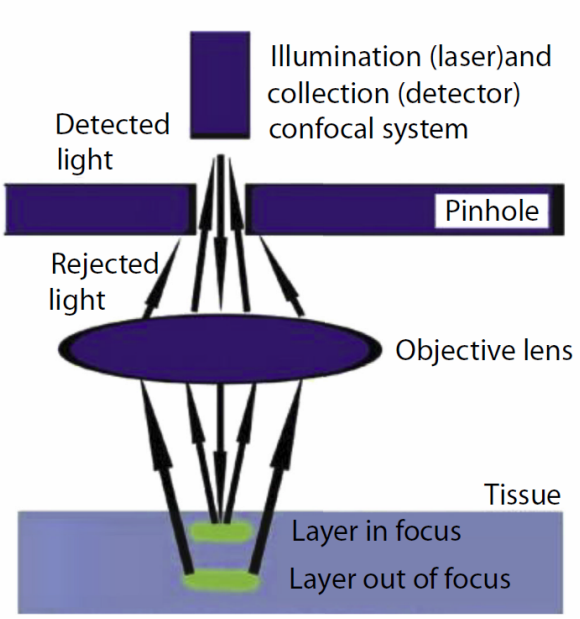

Figure 10: An explicative scheme of the inner workings of a confocal endoscope.

iology of the vasculature in a manner relatively unfamiliar to pathologists: it shows vessel density and shapes, it accumulates near leaky capillaries, and it accumulates in the lumen in areas of ulceration. Like eosin, it provides no direct visualization of nuclei, though the relative absence of fluorescein in the nuclei can allow some estimate of nuclear size and position [60; 62; 63; 64]. For nuclear staining, 0.05% topical acriflavine hydrochloride (Sigma Aldrich, St Louis, Missouri, USA) [64; 65] can be used. Acriflavine was originally used in the early 20th century as an antimalarial and antibacterial drug, with little evidence of toxicity in humans [66]. Acriflavine, though, binds to nucleic acids, and is potentially mutagenic [62]. Nevertheless, several studies in humans have documented the usefulness of topical acriflavine in GI endoscopy [64; 65; 67; 68]. Besides concerns about mutagenicity, acriflavine has limited penetration into tissue, and gives relatively uneven staining [63]. Practically, patient preparation for CLE is the same as preparation for conventional endoscopy. The procedure is then performed under moderate sedation or general anesthesia depending of patient co-morbidities and the type of procedure (i.e. colonoscopy, upper endoscopy, endoscopic retrograde cholangiopancreatography or endoscopic ultrasound). In addition to the contraindications for conventional endoscopies, CLE also has the additional contraindication of allergy to contrast agents. CLE can be applied in GI endoscopy in Barrett's Esophagus surveillance and treatment, diagnosis of biliary structures (pCLE), follow-up of colonic regions, differentiation of colorectal polyps, IBDs and gastric diseases. A novel reusable probe (needle-based CLE, nCLE) that is designed to advance through a 19-gauge needle is now available and could be used for analysis of pancreatic cysts. Typically, gray scale images are collected from CLE analysis. Images from adjacent planes could be blended to create 3D images [62], but commercial tools for this application do not exist yet.

2.5 Computer aided decision support systems in flexible endoscopy

The advent of endoscopes with the ability to take digital pictures created the whole new field of computer-aided decision support systems (CADSSs) in medical endoscopy. Such systems are designed to detect and/or classify abnormalities and thus assist a medical expert in improving the accuracy of medical diagnosis. In addition, different methods have emerged which do not directly provide decision-support. Instead they aim, for example, at enhancing image quality, detecting degraded images, or provide endoscope navigation support. Throughout this chapter, such methods are referred to as supportive systems. There is a rising interest in this research topic, starting about one decade ago [69].

A rough overview of common steps involved in a decision support system for medical endoscopy has already been shown in Fig.1. In many cases the first step is a preparation of the tissue region to be investigated (e.g., staining, treatment with fluorescent dyes). After an image has been acquired, preprocessing may be required in order to enhance the quality of possibly degraded images. Then, depending on the aim of the application, suitable features have to be found and extracted. Sometimes a post-processing of the features is also necessary (e.g., removing invalid feature combinations in the case of high-level features). If the decision support system is targeted at classification (e.g., polyp detection, cancer detection) the features are used for a classification of the image, using a previously trained classifier. But other systems also exist which base their decisions directly on the features without using an intermediate classifier (e.g., by using feature thresholds) [70]. Similar to classification, some systems are targeted at content-based image retrieval (CBIR) or content-based video retrieval (CBVR). The main difference between automated decision support systems and CBIR/CBVR systems is the fact that, in case of an automated decision support, the output of such a system is a suggestion on the final diagnosis or additional information for a diagnosis. This output is usually generated without any intervention by a medical expert being needed, potentially allowing, for example, a real-

time polyp detection while the endoscope is advanced through the colon. CBIR/CBVR systems on the other hand present an expert a number of similar images or videos (on demand), from which the expert is able to decide by himself on the final diagnosis. It is also quite common that the expert is able to interact with the system, allowing a refinement of the search query for similar images. Hence, CBIR/CBVR systems usually have an interactive nature which limits them in terms of real-time capabilities and restricts them to offline processing usage.

As already pointed out above in Sect. 2.4, each endoscopic procedure generates images which exhibit peculiar characteristics depending on the technique used. Therefore, computer systems targeted at decision support must be designed accordingly. Images taken with a traditional flexible endoscope don't allow us to see details of the tissue under examination. A zoom-endoscope, on the other hand, allows one to examine the fine structures and details of tissue too. This, however, comes along with a rather limited field-of-view (FOV), which makes navigation more difficult. This problem is even more apparent in the case of CLE due to the high magnification nature of this technique, that can although produce images which contain clear and detailed structures. As it is pointed out in Sect. 2.4, the distance between the tip of the endoscope and the mucosa under inspection differs between these techniques. This is due to the different focal depths inherent to the different techniques. As a result, the FOV differs also between the devices. While standard endoscopes usually have FOVs between 120° and 170°, zoom-endoscopes have rather limited FOVs between 50° and 70°. This naturally affects the size of the visible mucosa regions. In case of CLE the FOV is even more limited, resulting in a visible region of about 500×500 µm. Nevertheless, the limited FOV comes along with the advantage of higher image resolutions. In the case of WCE, the image resolution is often considerably lower compared to the aforementioned techniques. In addition, WCE suffers from the inability to control the motion and position of the capsule, which raises new difficulties for CADSSs. It is clear that, even in case of the same pathology, images taken with different endoscopic techniques will in general differ significantly. One particular difference

between the different endoscopic modalities is the available image resolution: these range from approximately 0.65 Megapixel to approximately 2 Megapixel. In addition, while some endoscopes allow us to capture color images, there also exist endoscopes which capture grayscale images only.

Since this chapter's main point is to present facts and figures for CADSSs in the GI tract, the spread of the different endoscopic imaging modalities across CADSS-related literature will be discussed next. For this purpose an overview of the different parts of the GI tract for which CADSSs have been developed in the past is given first. Then the different pathologies under investigation are outlined, showing the importance of respective detection and classification systems. Finally, different approaches found in literature from the image processing and classification perspective are shown, providing details such as used transformations, features and classifiers. WCE has a greater need for CADSSs then WLE, given the huge number of images it generates. Since the analysis of all these images by a medical expert is a time consuming task, it is a logical consequence that there is a rising interest in developing CADSSs for WCE. Because of the fact that CLE is the most recent technique, the number of respective CADSSs targeting this technique is still low. At the time of this writing, the methods in the literature are all based on pCLE. As of today, there exists no CADSSs related work based on eCLE in the literature.

In 2011, Liedlgruber and Uhl [69] showed the number of publications found in literature dealing with CADSSs using different endoscopic imaging modalities. This review has shown that flexible endoscopy is clearly the most frequently targeted endoscopic technique (about 58%), followed by WCE (about 38%), and pCLE (about 4%). Concerning the sector of the GI tract under analysis for CADSSs development, the vast majority of presented CADSSs are aimed at the colon, given that colon cancer is the third most common malignant disease in western countries. These are based on flexible endoscopy and pCLE. 29% of all CADSSs in the GI tract aim at a complete analysis of the whole GI tract, usually with WCE. The most common diseases under analysis with CADSS are, in or-

der, colon polyps, bleeding of the GI tract, normal/abnormal classification of WCE images or colon images, ulcers, celiac disease and tumors in the small bowel using flexible endoscopy. Regarding features, the most commonly used are:

- Spatial domain features:
  - Texture Properties: Local Binary Patterns (LBP) and its extensions (uniform LBP, Opponent color LBP), Texture Spectrum (TS);
  - Pixel-based: pixel intensities or other statistics;
  - Histograms: Color histograms or co-occurrence histograms from which statistical measures are computed;
  - Miscellaneous: SIFT, run-length features, MPEG-7 descriptors, blob analysis, Bag of visual words, etc.
- Frequency domain features:
  - Fourier based: FFTs are used to obtain power spectrum of an image, from which statistical features are computed and used for classification;
  - Wavelets: DWT, Dual tree complex WT, Curvelet transform or Gabor wavelets are applied on the image. Then, features are extracted (such as statistical features, Haralick features in a subband, random field parameter estimates, histograms computed from LBP-transformed subband, shape parameters of probability distributions fitted to the wavelet coefficients).
- High-level features:
  - Edge-Based: Canny or SUSAN approaches, either using the edges directly as features or calculating shape descriptors on them;
  - Region-based: relying on some previous sort of segmentation or region growing. Features extracted from each region are

computed. Other approaches could use the regions to deduce relationships among them (based on distances, for example, among region centers or similar approaches).

Ground truth can be either histologically verified or only obtained visually, and validation can be done with cross validations (either with leave-one-image-out or with leave-one-patient-out) or simply with distinct sets for training, validation and testing. Classification is usually done with one of these methods (or an ensemble of them): support vector machines, k-nearest neighbors, artificial neural networks or related classifiers, Gaussian mixture models or discriminant analysis based classifiers. For each classification method used in this work, it will be thoroughly described in a dedicated section.

Chapter 3

CADSSs in Confocal Laser Endomicroscopy

To be able to extract quantitative measurements from images, it is necessary to extract meaningful features from each of the images. In the following sections, different processes and algorithms will be described, and each one of those will be based upon predefined features that need to be calculated from the images. The proposed CADSS has the following strengths:

- Villi detection and segmentation: two different methods will be presented that aim at creating a villi segmentation binary mask for each image in the dataset. The first process will be semiautomatic, and based on morphological processing. The second one is a fully automatic process, based on SLIC superpixel segmentation and machine learning.
- Goblet cells detection: two different algorithms will be presented: the first is based on matched filtering, morphological processing and Voronoi's segmentation, while the second method is a simpler and faster version, purely based on matched filtering.
- Crypts detection: a method to detect crypts in the images is briefly presented in this chapter, and thoroughly analyzed in Chapter 4.

- Leakage detection: a simple algorithm based on thresholding detects regions of the image exhibiting saturated intensities. Saturation is caused by leakage of the fluorescein contrast agent in the extracellular intervillous space, due to ruptures of the cells building the villous mucosa.

All of this will be implemented in a user-friendly GUI, presented in Chapter 4.
In the literature [71; 72; 73; 74; 75; 76; 77] several features have been proposed, that highly correlates with pathological conditions of the gastrointestinal mucosa. Among them, for celiac disease, there are architectural changes of villi and/or crypts, increased intraepithelial lymphocytes (IEL), increase in lamina propria cell density, reduced goblet cells' density. For increased intestinal permeability, instead, three features have been defined: Cell Drop-out (CDO), shedding of an enterocyte into the luminal space; Cell Junction Enhancement (CJE), a fluorescein build-up between two epithelial cells representing impaired tight-junction proteins before breakage of the final basal junction, and FL: a fluorescein plume entering the lumen representing loss of apposition between two adjacent cells. Images originating from CLE, as Fig. 11 shows, are very informative about the status of small bowel mucosa: villi, crypts and goblet cells can be clearly discerned, depending on the region of the mucosa under analysis and the staging of the patient. Experienced endoscopists can diagnose villous atrophy, crypt hyperplasia, fluorescein leakage or cell drop-out and stage intraephitelial lymphocytes with high accuracy using this imaging modality. As such, automatic methods for the detection and segmentation of informative patches of the images, for establishing goblet cells' and crypt's density, for understanding the severity of villous atrophy, for checking for fluorescein leakage or other problems in epithelial tight junctions are clinically very important, and a path that is still unexplored. With this thesis' work, a first analysis of this setting is performed, and some automatic algorithms are proposed to start tackling this field of medical imaging.

To our knowledge, in fact, no previous work has been published on image processing methods with the purpose of automatically classify

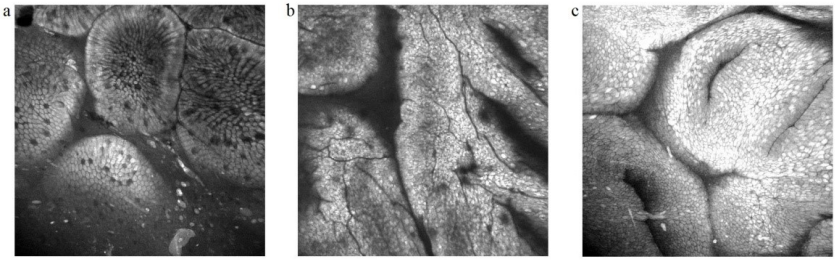

Figure 11: Three images from the dataset.

and stage celiac disease and other diseases from CLE images. This is not a trivial process: villi are highly textured and present high variability in appearance, shape and dimension. In CLE images, they can exhibit smooth and fuzzy borders among (and between) villi and inter-villous space. Among villi, vessels can be found in inter-villous space, with goblet cells. In severe CD stages, a possible collapse of all villi into a uniform mucosa that is depleted of villi can be observed. Other than this, possible presence of crypts and fluorescine leakage in the lumen (possibly caused by alterations in intestinal permeability) can prevent accurate detection with standard image processing methods. Normally (Fig. 11, left), villi are distinguishable one from the other. If Villous Atrophy is present (Fig. 11, center), the mucosa is flattened, resulting in villi disappearing in the focal plane. When Crypt Hypertrophy is present, the mucosa flattens on the top of the villi and cells' volume increase, resulting in bridges among different villi and more visible crypts (Fig. 11, right). Automatic grading of such problems could result in in-vivo virtual biopsies or in targeted biopsies (instead of using the four-quadrant protocol), decreasing the need for potentially dangerous multiple biopsies to grade the disease.

3.1 Semiautomatic villi detection via morphological processing

3.1.1 Materials

In this study, 166 confocal images were obtained from previous clinical trials conducted at the Gastroenterology and Liver Services of the Bankstown-Lidcombe Hospital (Sydney, Australia) [78]. Each patient underwent a confocal gastroscopy (Pentax EC-3870FK, Pentax, Tokyo, Japan) under conscious sedation and with a intra-venous aliquots of fluorescein sodium and topical acriflavine hydrochloride to enhance images. Each image represents a mucosal region of 0.5×0.5 mm, with an in-plane resolution of 2 pixel $/\mu\mathrm{m}$, resulting in images of 1024×1024 pixels. As Fig. 12 and 13 show, images conveying very heterogeneous information have been selected for the study presented in this section, for generalization purposes. 100 images have been used for parameter selection, and the remaining 66 images have been used to test the performances of the proposed method. In order to provide a ground truth, all images have been manually analysed by an expert, providing both an outline of each visible villus in the image, and and a seed point for each identified villus. In the 66 test images, an average of 4.86 (standard deviation: 4.49, median: 3.50) villi are shown in each image.

The first step in the proposed method aims at the construction of a rough segmentation identifying a candidate region of the image with the highest possibility of being part of a villous fold for each of the manual seeds. This is performed by filtering the image using a pipeline of morphological operations, designed to smooth the internal textured region of each villus. The segmentation is then obtained applying a quadtree decomposition (a simple and fast method that provides a first rough segmentation) of the resulting smoothed image. The resulting detection of candidate villi may result in multiple seeds belonging to the same segmented region (two separate villi may be merged by this step). In order to tackle this, a splitting procedure is applied to any candidate region where multiple different seeds are present. The split is based on both

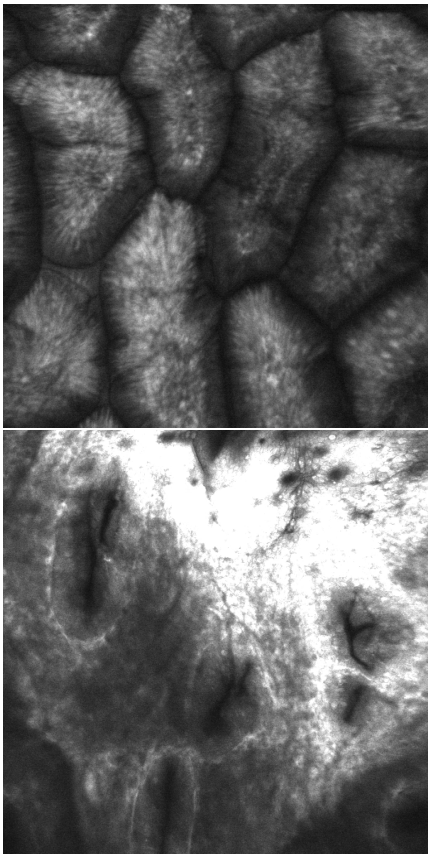

Figure 12: Two images from the dataset, showing very heterogeneous structures and illumination.

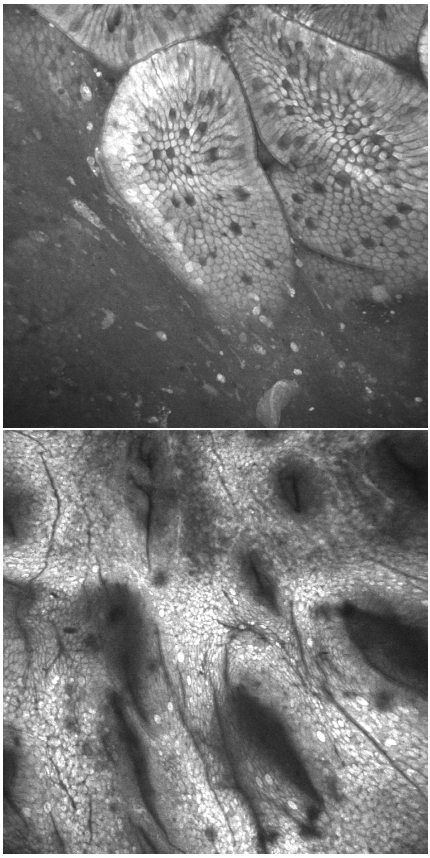

Figure 13: Two different images from the dataset.

local characteristics and shape of each candidate region hosting multiple seeds. After this process, each seed originates a different candidate region, and each of the regions hosts only one seed. The final processing step performs a refinement of the borders of each detected villus, so to obtain a final and accurate shape for each villus exploiting the local grayscale intensities of the original image.

3.1.2 Illumination correction

In order to correct for heterogeneous illumination in the images, a top-hat filtering has been applied to each image I as first step in a morphological processing pipeline of four different steps. In this process, a morphological opening (an erosion followed by a dilation) using as structural element a disk of radius of R_1 is applied to the input image I. The image I_o computed as such carries the information of the background illumination. The illumination-corrected image I_{ill} is simply calculated by subtracting the estimated I_o from the original image: $I_{ill}(x, y) = I(x, y) - I_o(x, y)$ with $x \in \{1, 2, \ldots, 1024\}$ and $y \in \{1, 2, \ldots, 1024\}$ being the coordinates of each pixel within the image.

3.1.3 Villi de-texture

As shown in Figs. 12 and 13 the shape and the appearance of the mucosa in each region of the image is very heterogeneous:

- goblet cells can be either clearly visible (Fig. 13, top) or out of focus (Fig. 12, top) in distinguishable villi;
- crypts can be the largest objects shown (Fig. 13, bottom);
- no sign of goblet cells and no clear border of the villus (Fig. 12, bottom).

Villi de-texture is performed through a three steps pipeline, composed by a series of morphological reconstructions with the purpose of reducing

the internal part of each villus of each image to a connected homogeneous set of pixels. Morphological reconstruction [79] among two images, the marker and the mask, consists of repeated operations of dilation of the marker image, until its contours fit the ones of the mask. The main result of this process is that the peaks of the marker image are dilated, creating homogeneous regions that originate from them. As the goal is to create smooth regions inside the villi (where the seeds lie), the first morphological reconstruction is performed using as the marker a rescaled version of the image, I_s, with its grayscale intensities reduced by an image-adaptive value c, and I_{ill} as mask, to obtain I_1. In order to provide a reconstruction whose behaviour adapts to the specific appearance of the villi in each image, the gray level intensity range r_i in a neighborhood of each seed $s_i(x, y)$ is computed, with the position of the i^{th} seed $s_i(x, y)$, $i = \ldots, N$, and N the number of seeds of the image. The average value among all seeds is retained as $c = 1/n \sum_i r_i$. The second and third morphological reconstructions are performed in series, to obtain I_2 and I_3, in which regions are progressively smoothed. They use as markers an eroded version of I_1 and a dilated version of I_2 respectively, and I_1 and I_2 as masks. The markers are obtained using disk shaped structural elements of radius R_2 and R_3 respectively. As Figs. 14 and 15 show, each of these steps ensures to smooth the image in the proximity of each seed.

3.1.4 Quadtree decomposition

Quadtree decomposition [80] is a well-known technique that is used in image processing to subdivide square images into four blocks, when a certain condition is met. Usually the condition involves the definition of a homogeneity criterion: as such, the decomposition adapts to the structure that each image is carrying. The final result can be a set of blocks of different size. In this work, the criterion is a threshold T for the range of the intensity values for each block. For each seed, the average grayscale intensity μ_i is computed in a circular neighborhood of $s_i(x, y)$ of radius R_4. Starting from the block B containing the seed, the connected set of

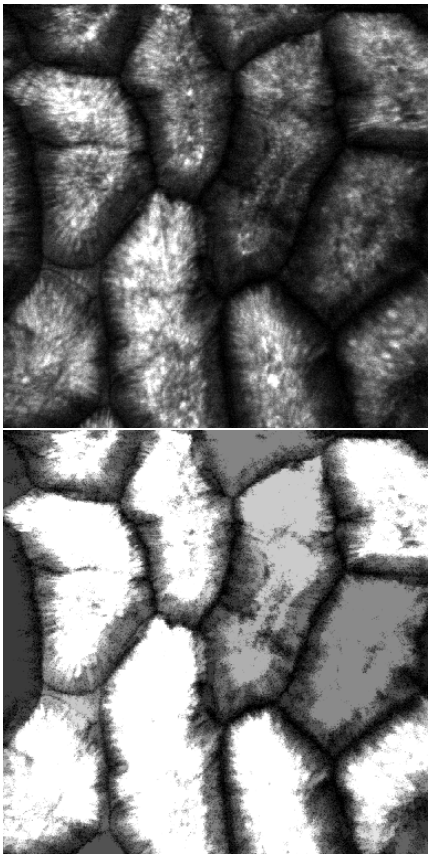

Figure 14: Processed images visually explaining the morphological processing pipeline steps: (Top) I_{ill} (Bottom) I_1

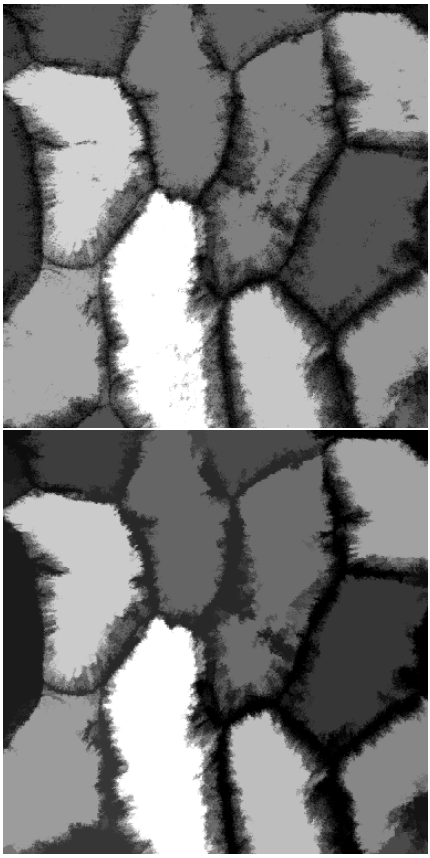

Figure 15: (Top) I_2 (Bottom) I_3

blocks resulting from the quadtree decomposition with average intensity μ_B for which the condition $|\mu_B - \mu_i| \leq \theta\mu_i$ is satisfied are assigned to the region V_i belonging to the i^{th} seed.

3.1.5 Splitting process

Since two seeds can collapse into the same region after the quadtree decomposition, a splitting process is required to ensure that each seed originates only one villus, and that one villus is defined by only one seed. If a region V contains S different seeds $s_i(x, y)|(x, y) \in V$, the possible splits are first detected considering every combination of possible different pairs of seeds $[S_i, S_j] : i \neq ji, j \leq S$. A segment $l_k, k = 1...i \cdot j$ is traced between $[S_i, S_j]$, and the pixel $P_k \in l_{ij}$ is selected as the pixels among those in l_k with the minimum Euclidean distance to ∂V, where ∂V is the border of the region V. The region V is then iteratively split considering all P_k splitting points ordered with increasing distance from ∂V. The split is obtained by cutting the region V with a line normal to l_k and passing through P_k. At the end of this process, only regions containing one seed are kept, and the selection of candidate villi $\{V_i, i = 1, \ldots, N\}$ concludes.

3.1.6 Final border refinement

Since the output of morphological processing and the splitting process only represents a rough estimation of villi's shape (see Fig. 16), a refinement process is needed for a reliable estimate of the borders of each villus. An operation of morphological closing with a disk shaped structural element of radius R_5 is performed to smooth the borders. Given a region V_i, and a point $Q(x, y) \in \partial V_i$, a line of length $L = 80$ pixels considered from $Q(x, y)$ and normal to ∂V_i. A new border point $Q'(x, y)$ is estimated as the point with minimum intensity along the line to define the new border $\partial V'$. Finally, third-order Savitzky-Golay smoothing filter [81] with window size $f = \min(25, l)$, where l is the length of $\partial V_i'$ contour. Overlap among different regions is avoided by imposing a minimum distance among each smoothed contour.

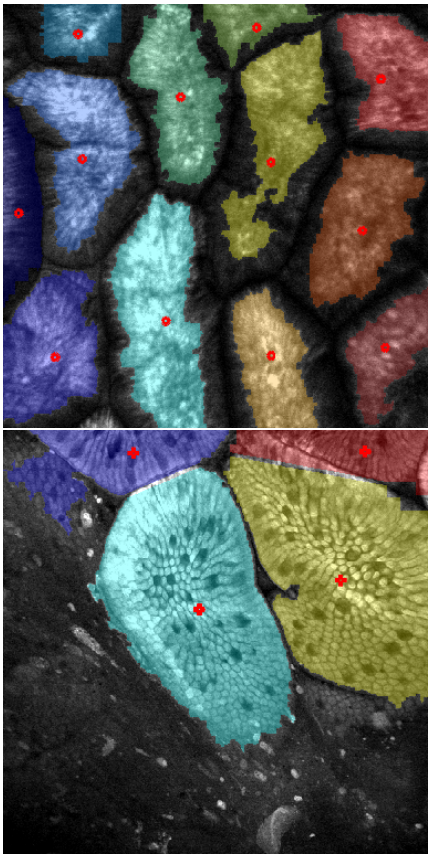

Figure 16: The first-step rough segmentation results superimposed on two images.

3.1.7 Results and conclusions

To quantify the performance of the proposed method, the Dice coefficient for each villus is computed by comparing the segmentation results with the respective manual ground truth. The values of $R_1 = 200, R_2 = 60, R_3 = 25, R_4 = 2, R_5 = 19, T = 0.08, \theta = 0.05$ have been heuristically set after analyzing the performance on 100 images among the 166 of the database. The proposed method has been tested on 290 villi, and reached an average accuracy of 81%. The median value of the accuracy among all 290 villi is 88%. Only in 22 cases the single accuracy scored below 50%. Figs. 17, 18, 19, 20, show four images from the dataset and the true and detected villi for visual comparison.

3.2 Automatic villi detection via superpixel segmentation

3.2.1 Materials

In this subsection, a different approach to the same problem is proposed. Here, the process is aimed at full automatic detection by using superpixel segmentation. For this, 155 confocal images were obtained from a previous clinical trial conducted at the Gastroenterology and Liver Services of the Bankstown-Lidcombe Hospital (Sydney, Australia) [73]. These image maintain the resolution of the image already presented (0.5×0.5 mm per image, in-plane resolution of 2 pixel/ μm, 1024×1024 pixel dimensions). Again, as shown in Fig. 21, images convey very heterogeneous information. Among the three CLE features (fluorescein leakage, cell drop-out and cell junction enhancement), each image of the dataset exhibit only one feature. In total, the dataset is composed by 29 CDO images, 65 CJE images and 61 FL images. A random selection of 70 images has been used for populating the training set. Another random selection of 15 of the remaining images was used to tune the post-processing analysis, a step that will be detailed in the following subsection. The remaining 70 images were used for testing the method's performance. All images

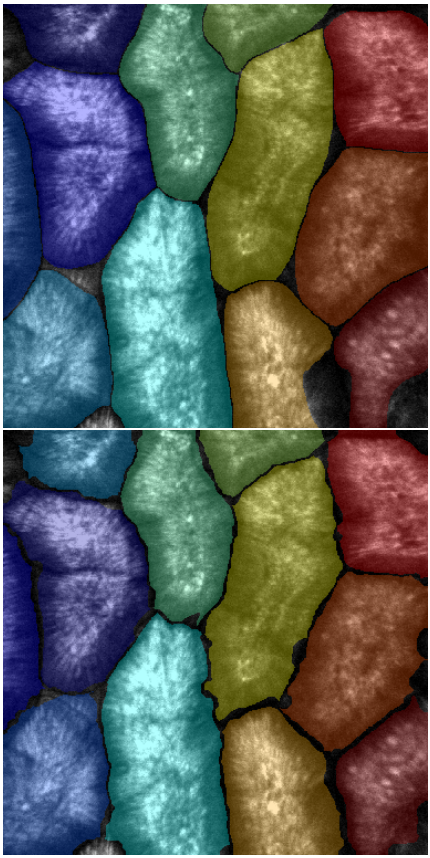

Figure 17: (top) ground truth labeling and (bottom) villi estimated area superimposed on an image from the dataset. The algorithm performs almost perfectly on this image (mean Dice coefficient is 94%).

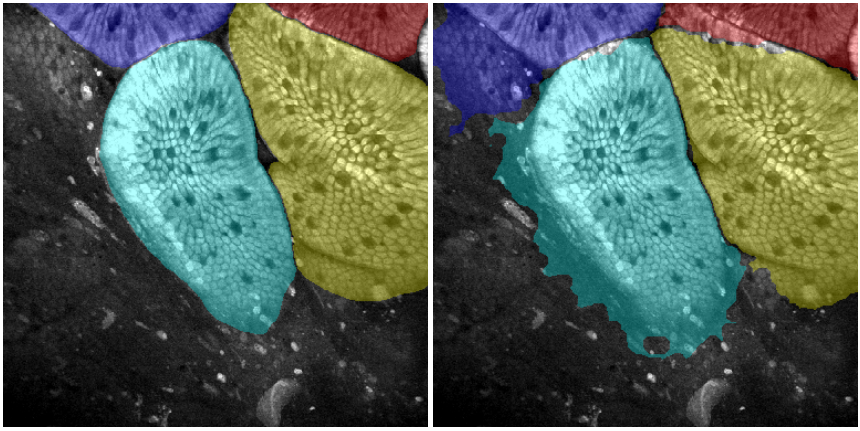

Figure 18: (left) ground truth labeling and (right) villi estimated area superimposed on an image from the dataset. The algorithm performs almost perfectly on this image (mean Dice coefficient is 98%).

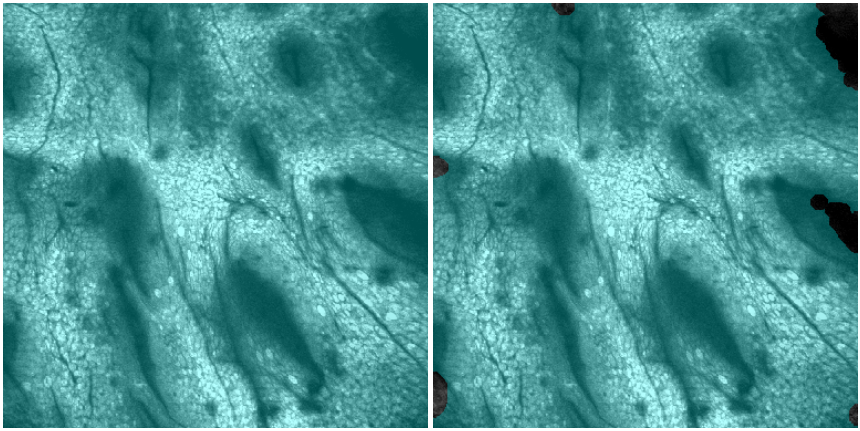

Figure 19: (left) ground truth labeling and (right) villi estimated area superimposed on an image from the dataset. The algorithm performs almost perfectly on this image (mean Dice coefficient is 99%).

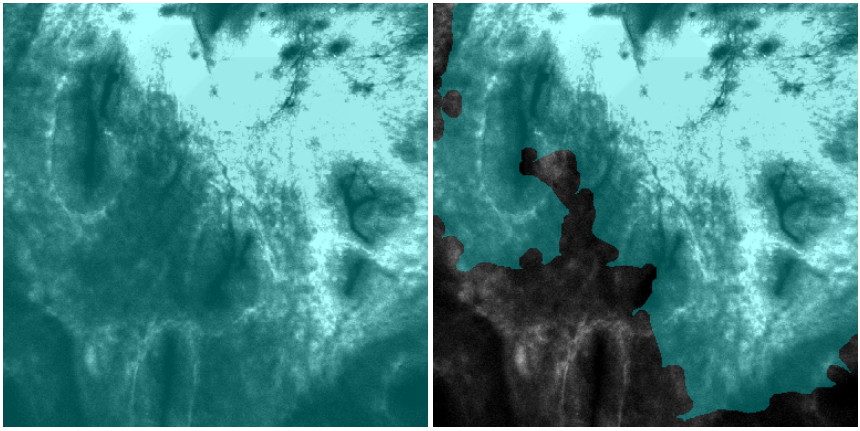

Figure 20: (left) ground truth labeling and (right) villi estimated area superimposed on an image from the dataset. The algorithm encounters difficulties with this image (mean Dice coefficients is 83%).

have been manually analyzed by an expert, providing an outline of each visible villus in the image as a ground truth.

The first step in the proposed method aims again at the construction of a rough segmentation identifying a candidate region of the image with the highest possibility of being part of a villous fold. This is performed by processing the image with a computer vision technique called superpixel segmentation, in particular using the SLIC implementation [82]. The purpose of this process is to create clusters of spatially connected pixels exhibiting similar texture. Each of the superpixels is then analyzed, and 37 features are extracted from each of them, to be fed to a classifier. A multiscale analysis is performed, by computing and analyzing three versions of the original image (original size, plus two rescaled versions by a factor of $1/2$ and $1/4$ respectively), bringing the total size of the feature vector for each superpixel to 111. The classification step is performed with an ensemble of 50 decision trees, trained on 70 random images from the dataset. A post-processing refinement step of the computed prediction is performed to improve the accuracy, tuned on a sub-sample of the image dataset (15 images, referred to as "tuning set") to maximize ground truth

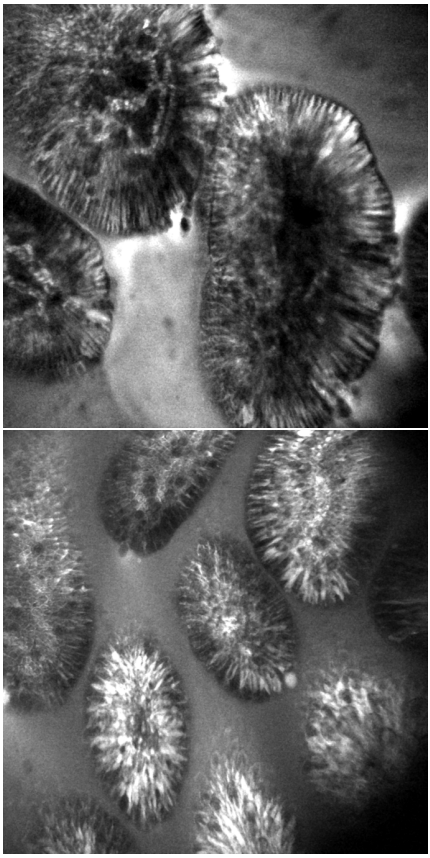

Figure 21: Two images from the dataset.

adherence and prediction accuracy. The algorithm is then tested on the remaining 70 images.

3.2.2 Superpixel segmentation

As a pre-processing step for each image, all greyscale values were normalized between 0 and 256, and a median filter (of size 3×3) was then applied to reduce noise. Segmentation via superpixel is then performed by grouping pixels into perceptually meaningful atomic regions, used to replace the rigid structure of the pixel grid. Many computer vision algorithms use superpixels as their building blocks [83; 84], given their straightforwardness and the ease of their implementation. A commonly used superpixel implementation is the Simple Linear Iterative Clustering (SLIC) [82]: this implementation, based on k-means clustering, is fast to compute, memory efficient, simple to use, and outputs superpixels that adhere well to image boundaries. SLIC was designed with the following properties in mind:

- Superpixels should adhere well to image boundaries;
- Superpixels should be fast to compute, memory efficient and simple to use;
- Superpixels should both increase the speed of segmentation and improve the quality of the results.

This superpixel generation algorithm is distinct from k-means clustering in two important ways:

- the number of distance calculations in the optimization step is dramatically reduced by limiting the search space to a region proportional to the superpixel size. This reduces complexity to be linear in the number of pixels N (and independent of the number of superpixels k);
- A weighted distance measure combines color and spatial proximity, while providing control over the size and compactness of the superpixels.

SLIC clusters pixels and efficiently generate compact and nearly uniform superpixels, imposing a degree of spatial regularization to extracted regions. This technique only requires two parameters to set: the desired size of each superpixel N and a regularization parameter λ, that tweaks the smoothness of their contours. The entire algorithm is summarized in Algorithm 1.

/* Initialization*/;
Initialize cluster centers $C_k = [l_k, a_k, b_k, x_k, y_k]^T$ by sampling pixels at regular grid steps S.;
Move cluster centers to the lowest gradient position in a 3×3 neighborhood.;
Set label $l(i) = -1$ for each pixel i.;
Set distance $d(i) = \inf$ for each pixel i.;
repeat
  for *each cluster center* C_k **do**
    for *each pixel* i *in a* $2S \times 2S$ *region around* C_k **do**
      Compute the distance D between C_k and i.;
      if $D < d(i)$ **then**
        set $d(i) = D$;
        set $l(i) = k$;
      end
    end
  end
  /* Update */;
  Compute new cluster centers;
  Compute residual error E.
until $E \leq$ *threshold*;

Algorithm 1: SLIC Superpixel segmentation.

SLIC superpixels generally correspond to clusters in the $labxy$ color-image plane space. While in this explanation LAB color space will be the reference, remember that all the images used in this work are monochromatic. The measure D defines the distance between a pixel i and a cluster center C_k, as defined in Algorithm 1. Color proximity and spatial proximity are normalized to their respective maximum distances within a cluster, N_s and N_c. While the maximum spatial distance within a

given cluster should correspond to the sampling interval, color distances can vary significantly from cluster to cluster, and from image to image. Therefore, the maximum color expected is fixed to a constant, that allows to weigh the relative importance between color similarity and spatial proximity. When the constant is small, the superpixels adhere tightly to image boundaries, but have less regular size and shape. About its complexity, by localizing the search in the clustering procedure, SLIC avoids performing thousands of redundant distance calculations. In practice, in fact, one pixel falls in the neighborhood of less than eight cluster centers, meaning that SLIC is $O(N)$ complex (while the upper bound for classical k-means clustering is $O(k^N)$.

Two typical images from the dataset, with superpixels superimposed, are shown in Fig. 22. This step has been implemented with MATLAB R2015b, using a third-party implementation of SLIC superpixels bundled in *vlfeat* [85]. Once the superpixel segmentation is obtained, the manual ground truth is transformed in superpixel space, as Fig. 23 illustrates. Each region of this image (corresponding to each computed superpixel) is labeled as part of a villous fold if, for that superpixel, the ratio among villous-labeled pixels and background-labeled pixels is greater than a threshold R, whose value is computed as explained in Sec. 3.2.5.

3.2.3 Feature extraction

For each image in the dataset, three different scales are analyzed for feature extraction: the image at the original scale, and two rescaled versions of it by factor of $1/2$ and $1/4$, respectively. In this way, a multiscale analysis of each image is performed, to improve robustness of the classification and to avoid possible errors due to texture similarities at the original scale. A total of 111 features are extracted for the multiscale analysis of each superpixel S, 37 for each image scale:

- Mean intensity μ_S and standard deviation σ_S: greyscale intensity variations can be useful features to differentiate among villous folds (i.e., foreground) and mucus (i.e., background);
- Contrast C_S, Energy E_S and Homogeneity H_S from the Gray Level

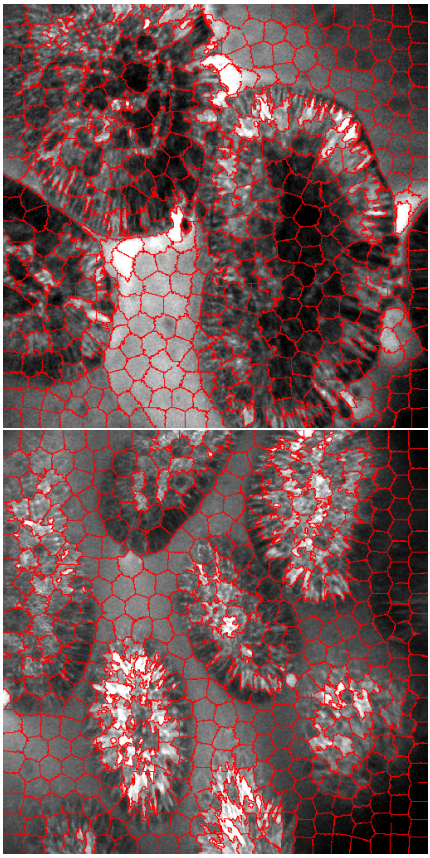

Figure 22: Two images with SLIC Superpixels superimposed.

Co-Occurrence Matrix (GLCM): GLCM is a statistical method of examining texture considering the spatial relationship of pixels. It calculates how often pairs of pixels with specified values and spatial locations occur in an image, building a 8×8 occurrence matrix. Extracting statistical measures from this matrix provides information about the specific texture. From this analysis, contrast (local variations in the GLCM), energy (sum of squared elements in GLCM) and homogeneity (how close the distribution of the elements in the GLCM is to its diagonal values) measures have been included in the feature set;

- Histogram of Local Binary Patterns [86] with 32 bins, $hLBP_S$. Local Binary Patterns (LBP) are one of the most descriptive features in the field of texture classification, and are commonly used in computer vision. They permit the creation of features able to identify different textures in an image. In this work, for each pixel of the image, an 8-bit word is created by comparing its greyscale intensity value with the ones in its 8-neighborhood. Iteratively, starting from a fixed direction, if the central pixel has a grayscale value greater than its neighbor a 1 is encoded in the 8-bit word, a 0 otherwise. When a word has been assigned to each pixel, each word is translated to decimal (0-256). A histogram (32 bins) is then computed for the LBP of pixels in each superpixel, expressing in such way the spectrum of the texture of the selected portion of the image. This is finally added to the feature vector.

3.2.4 Classification with random forests

For each superpixel, the probability of it being part of a villus fold is computed as the score of a binary random forest classifier using 50 classification trees. The training process has been performed using 30870 superpixels belonging to the 70 images from the training set.

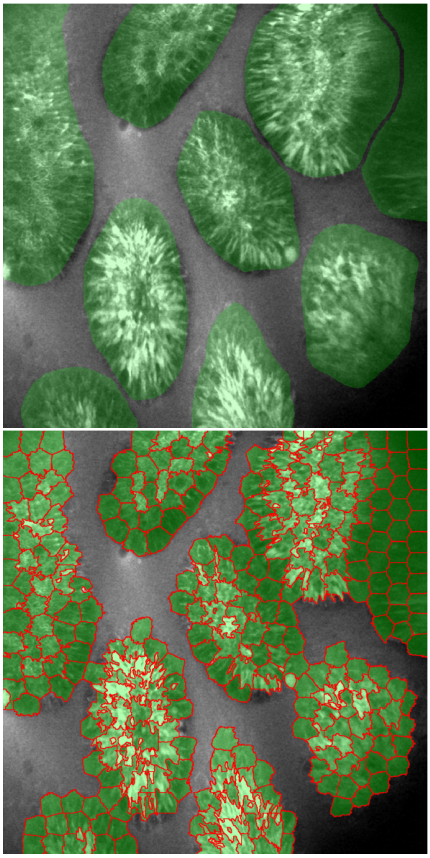

Figure 23: From manual ground truth (top) to superpixel-based ground truth (bottom).

3.2.5 Final refinement and results

After the classification step, binary prediction masks have been created according to the score assigned to each superpixel by the classifier. To discard isolated superpixels selected as villi, all connected regions smaller than P pixels have been excluded from the prediction masks, and holes in the binary masks were filled, to compensate for obvious false negatives in the classification step. The tuning of this final refinement process was based on the tuning set, composed by images excluded from both the training and the testing phase of the classifier. Accuracy, sensitivity and specificity of the classification step have been computed, both in superpixel and in pixel space, along with the Dice scores between the prediction masks and the superpixel based ground truth.

Superpixel parameters were set as $N = 50, \lambda = 0.05$ to obtain a number of about 440 superpixels per image, each of them resulting well adherent to image borders. The value of $P = 15742$ was tuned by selecting the maximum area (smaller than $P_{lim} = 16900$ pixels, as a hard-coded safety value based on villi's sizes from the tuning set, corresponding to a patch of 130×130 pixels) among all false positive villi identified in the tuning set. The value of $R = 0.5$ was set by maximizing Dice correlation among the labeled ground truth in pixel space and the one in superpixel space in the images of the training set (Dice score between pixel-space GT and superpixel-space GT at $R = 0.5$ is 96.3%). To quantify the performance of the proposed method, the Dice coefficient for each image is computed by comparing the prediction masks with the respective ground truth both in superpixel space and pixel space. The proposed method has been tested on 70 images (a total of 336 villi), and reached an average general accuracy of 85.9%. Sensitivity (True Positive Rate, TPR) is 92.9%, while specificity (True Negative Rate, TNR) is 77.0%. Mean Dice values between each prediction mask and its ground truth in superpixel space is of 87.4%, and the pixel-wise total classification accuracy in pixel space is 86.4%. Sensitivity and specificity referring to the pixel domain are, respectively, 93.50% and 71.59%. Fig. 24 shows two images from the dataset and the true (a-c) and detected (b-d) villi for visual comparison,

in superpixel space.

3.3 Goblet cells detection and density estimation - Morphological processing

Briefly, the motivation of highlighting goblet cells in CLE images is strong: goblet cells density in the duodenal epithelium is used as a complementary measure of CD severity together with villi appearance and crypt presence. In fact, a reduction of goblet cells density is among the architectural changes that occur in the pathological mucosa of an individual affected by celiac disease, along with VA and CH. With an automatic tool to measure such density, experts will be able to better score the severity of CH and VA exploiting this new information, with a significant impact in the medical practice. Hence, along with the evaluation of VA and CH, goblet cell density is an important marker of mucosa damage: the ultimate goal of this part of this thesis is to provide a system that, exploiting the correlation among the density of goblet cells in the columnar epithelium and the severity of VA and CH, can help the medical expert in accurately staging the disease, potentially reducing the number of performed biopsies.

The first step in the proposed method is cell detection: starting from the ROIs, this is obtained by matched filtering and regional maxima transform. A cell segmentation step via Voronoi diagrams defined an approximate shape for each cell; the mean intensity value in each of those was computed. The mean μ_b and standard deviation σ_b of the (mean) intensity of the cells whose centers has been obtained by the bright matched filter have then been evaluated. An images-specific threshold θ, used as a discriminative threshold to label the dark cells as goblet cells or other dark area, is then estimated as $\theta = \mu_b - A\sigma_b$, where A is to be intended as a reinforcing factor to exclude from the computation all the candidate cells that might be bright outliers. After estimating an image-specific threshold to distinguish between goblet cells and other dark areas, a border refinement method based on region growing is then applied to each identified goblet cell, so to improve the final segmentation results. From

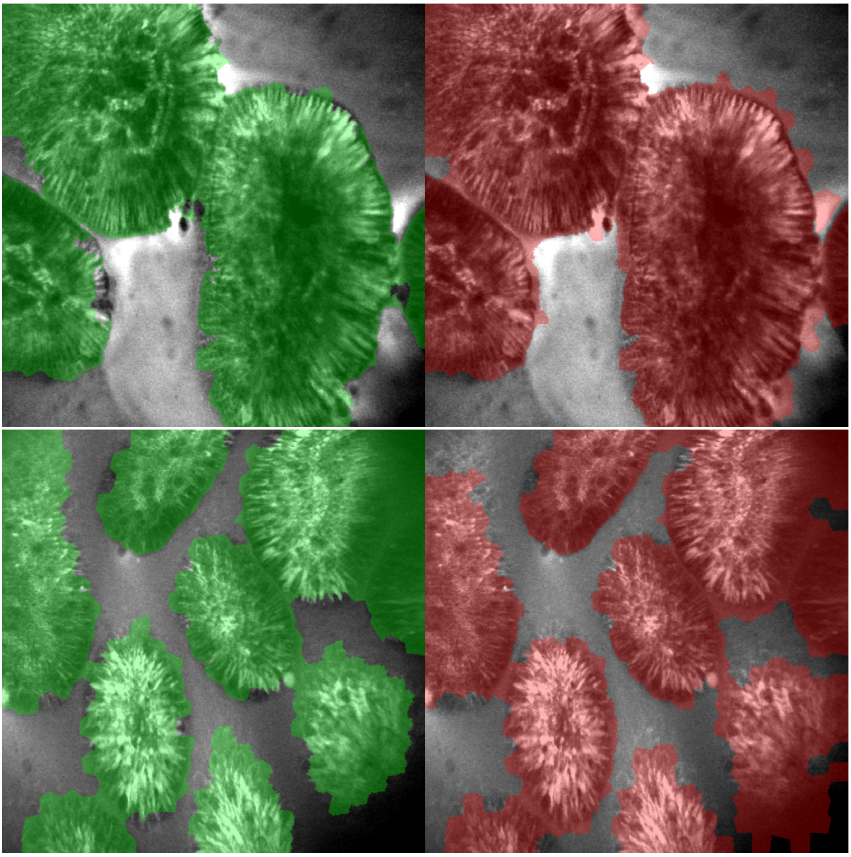

Figure 24: (a-c) ground truth labeling and (b-d) villi estimated area superimposed on two different images from the dataset.

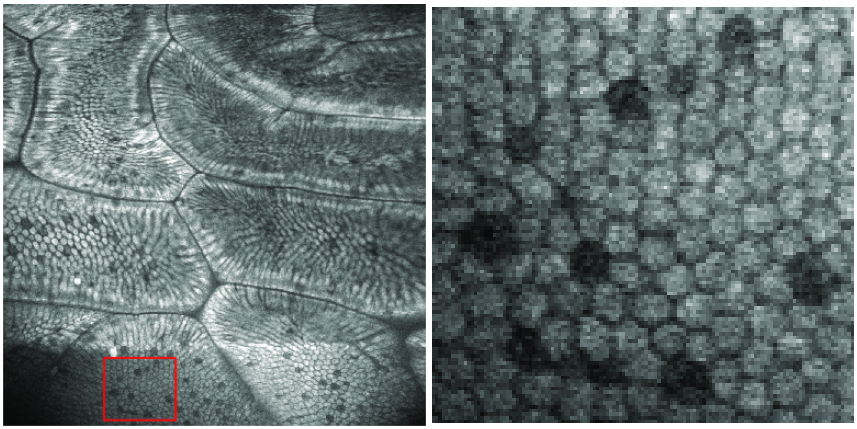

Figure 25: The process of selecting a ROI (right) from a full image (left), to lower computational complexity and to avoid inclusions of blood vessels, crypts or villi. Note how the ROI shows a majority of bright cells and a few dark goblet cells.

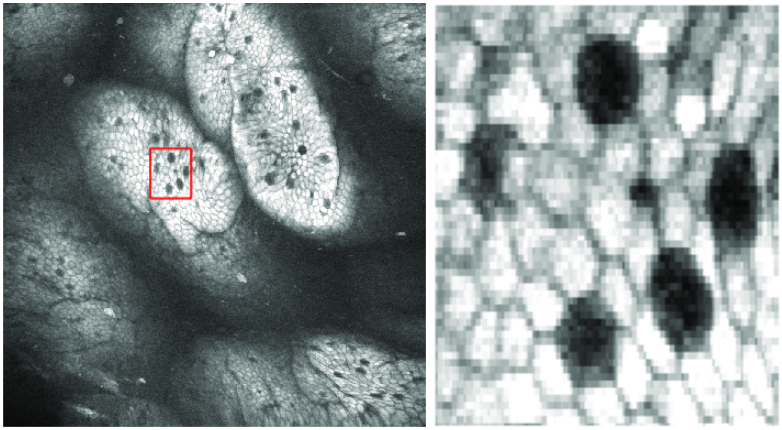

Figure 26: A ROI (right) from a full image (left).

the dataset, 45 images have been selected with the same characteristics as the ones already presented in the sections above, this time including only images showing goblet cells. As Figs. 25 and 26 show, a selection of the region of interests was needed because of the information contained in the full images: structures like blood vessels, crypts, artifacts, villi or intracellular dark tissue as well as epithelium not exhibiting goblet cells might cause misclassifications and penalize the automatic analysis. This step, in the work presented in this section, has been performed manually.

3.3.1 Preprocessing and matched filtering

In order to reduce the noise in the ROIs, a Wiener filtering [87] was applied to all the ROIs. All images were then normalized to have zero mean and a standard deviation of 1. As shown in Figs. 25 and 26, each ROI exhibits a majority of bright cells (from the duodenal columnar epithelium) and a minority of slightly bigger, darker cells (mucin-rich goblet cells). Two complementary circular matched filters (with radius of size comparable to the size of one cell, $r = 12$ pixels) have been designed with the purpose of enhancing all the cells in each ROI. To detect all the bright cells, a multiscale circular template has been used to define a filter bank, whose maximum response across the different scales enhances the bright cells (Fig. 27, middle) via convolution with each treated ROI. The regional maxima transform, which selects 8-connected component of pixels whose external boundary pixels all have a lower value, was then applied to the result of the convolution, to select the square regions of size $N \times N$ having altogether an intensity value greater than all their neighbors. These sets of pixels were selected as the centroids of the bright cells. In the same way, a second bi-dimensional filter was implemented, complementary to the first one (with negative values in its circular part), to isolate all the potential dark cells. As a second step, after the convolution (Fig. 27, bottom), the same regional maxima transform isolated and selected candidate centroids for goblet cells.

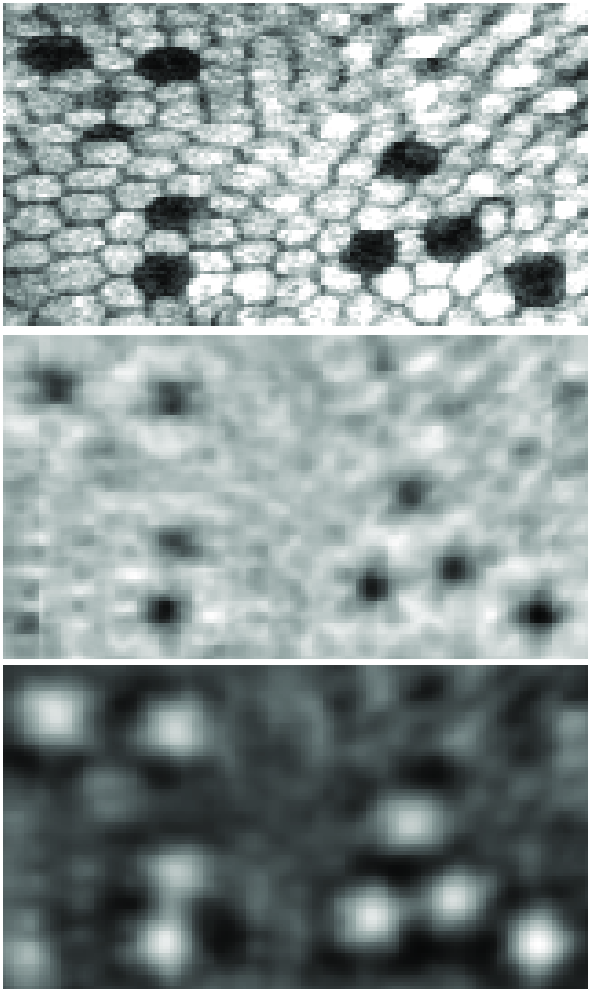

Figure 27: A ROI (top) and the images obtained through matched filtering using the bright (middle) and the dark (bottom) template.

3.3.2 Candidate cells detection

In all these $N \times N$ regions, mean, entropy and range were calculated. Given the particular heterogeneity in cell appearance within the epithelium (even within a selected ROI), not all dark spots are goblet cells, as can be appreciated in Fig. 27: in some of the ROIs, dark spots in the image represent intracellular space or thick structural borders, which need not to be erroneously labeled as goblet cells. To reduce the dark candidates and to minimize false detections, a selection process was applied to confirm as centroids only the $N \times N$ regions exhibiting mean intensity value $\mu_d < \theta_\mu$, entropy $E_d < \theta_E$ and intensity range $r_d < \theta_r$.

3.3.3 Voronoi diagram and mucin-cell selection

After the detection of the candidate bright and dark cells, the aim is identifying the cellular pattern, so to refine the cells' description corresponding to each centroid.
Using as centroids all the regional maxima obtained by the matched filtering processes, a Voronoi diagram [88] (using as a metric of closeness the Euclidean distance) was computed for each ROI, obtaining a partition of the image into regions, using a distance metric as a criterion to assign each pixel to a region or another. Using the centroids obtained by the matching filtering as the subset of points belonging to different regions, a rough cell segmentation of each image is obtained.
Each of the cells identified by the Voronoi diagram has then been eroded to exclude borders and possible interfering surrounding structures and has been characterized by means of its grey-scale mean value. Mean μ_b and standard deviation σ_b of all the average values of the cell originating from the bright matched filter computations were calculated. To identify the goblet cells among the candidates originating from the dark matched filtering computation, a thresholding process based on the mean intensity of each candidate cell has been performed: only the candidate cells with mean intensity value below the threshold $\theta = \mu_b - A\sigma_b$ were classified as goblet cells (i.e., only the significantly dark ones).

3.3.4 Refinement and classification

Since Voronoi's diagram is only a rough estimate of the real borders of each cell, it is sensitive to the presence of double positives, meaning multiple candidates resulting from multiple local maxima response from the matched filters that are close to each other, effectively belonging to one single cell. To tackle this problem, and in order to better segment each identified goblet cell, a region growing approach has been applied. For each detected goblet cell, a circular region centered on its centroid with radius equal to R_{min} is selected, and on this region the mean intensity value is computed. Then, the external border of the circular region is computed, and its average intensity value μ_i is calculated. If $\mu_i < \theta$, the pixel within the border whose value is lower than θ are added to the original region. The procedure is then iterated by increasing the value of the radius by one. When $\mu_i > \theta$ the procedure ends, and the final cell is computed as the convex hull of the resulting region. As final step, goblet density is estimated for each ROI, along with the percentage error and the correlation between real and estimated densities.

3.3.5 Results

True positives (TP) and false positives (FP) were computed for each image, comparing the detection and segmentation results to the ground truth. The value of θ was optimized by varying the variable A in $\theta = \mu_b - A\sigma_b$. The optimum value of A was found by using a receiver operating characteristics (ROC) analysis (Fig. 28) considering the true positive ratio (TPR, average on all images of the ratio among TP and the amount of real positives, according to the ground truth) and the number of FP per image. The optimal value of θ was found by selecting the farthest point of the ROI from the line drawn in the ROC space between its first and last points, in a similar manner to Youden index [89]. Results have been computed using a Leave-One-Image-Out technique, iteratively selecting $\hat{\theta}$ on 44 images and testing its value on the remaining one. The other parameters, after performing parameter tuning on a subset of the image database, were set using these values: $N = 2$, $\theta_\mu = 0$, $\theta_E = 0.1$,

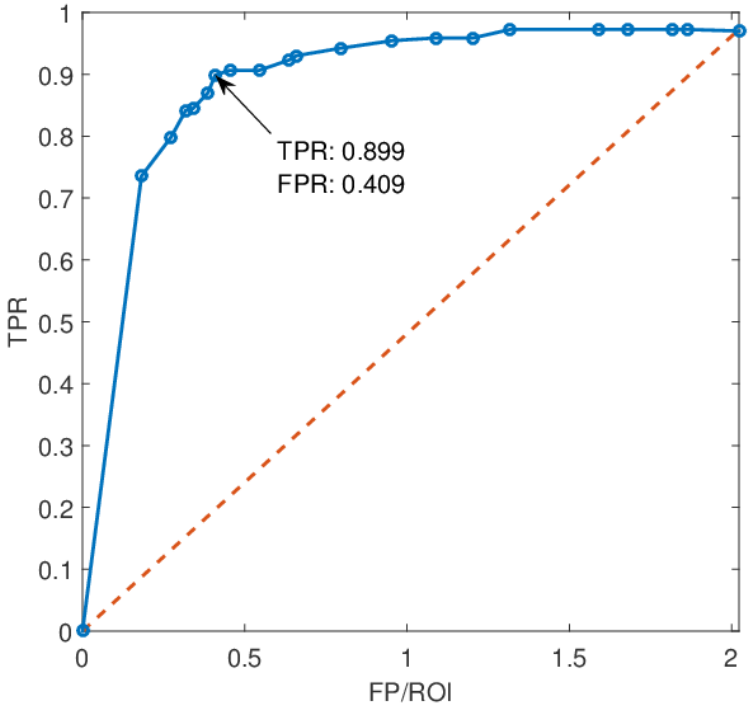

Figure 28: A ROC curve for the optimization of θ. The farthest point from the red line is selected as optimal value.

$R_{min} = 3$, and $\theta_r = 1$.

The proposed method reached 90.3% sensitivity with an average of 0.44 FP per image (0.22 ± 0.13s per image on a Intel i7-3630QM @ 2.40GHz, 8GB DDR3 RAM with an average image size is 130×94). Accepting one false detection per image (i.e., $FP = 1$), the sensitivity reaches 95.5%. The detection and segmentation of goblet cells in regions of interest is accurate and precise, as Fig. 29 shows.

The average estimated and average real goblet densities are comparable: the percentage error w.r.t. the real density is $9.7 \pm 16.9\%$, while

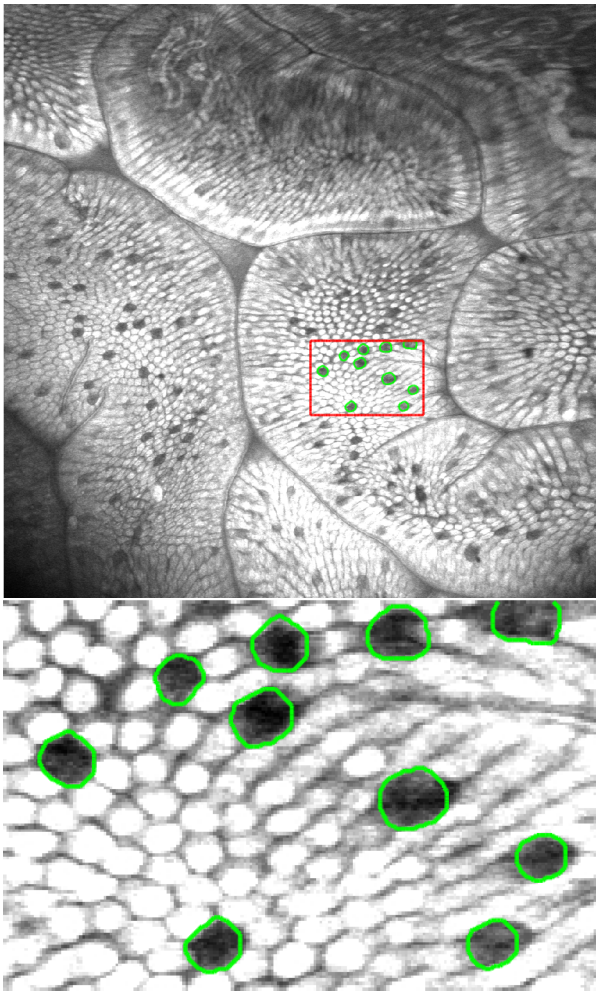

Figure 29: Results on a ROI (bottom) and superimposed on a full image (top)

the correlation is 87.2%. In Fig. 30, a scatter plot of the correlation among estimated and true densities is shown. The R squared coefficient of the regression line is $R^2 = 0.761$, meaning that 76.1% of the variance of the data is predicted by the linear fit.

3.4 Goblet cells detection and density estimation - Fast alternative method

A faster alternative way for goblet cells' and crypts detection has been implemented. The aim, in this implementation, is to achieve faster results by bypassing both the Voronoi step and the bright cells analysis. Based on the results of the matched filtering obtained by convolving the image with the filter designed to respond to dark patches (response depicted in Fig. 31), an absolute thresholding step has been devised to isolate only the regions with the highest response to the filter. Those regions have then been filtered by area, and labeled as goblet cells. This step does not use superpixels or the slow Voronoi cell detection, reaching an accurate estimation in a small amount of time if compared to the previously detailed method. This step will be better explained in the next chapter, dedicated to presenting the GUI.

3.5 Crypt detection

A fast algorithm, inspired by the fast goblet cells detection method described above, has been implemented to detect crypts in confocal images. Given the estimated villi segmentation, the darkest regions in the villous folds are selected as estimated crypts. First, an image is normalized to zero mean and unitary variance (Fig. 32).

Then, all grayscale intensities are normalized again to the intensity range exhibited by all pixels included in the villi segmentation mask. The edited rescaled image is then transformed to a binary mask, and a morphological operation is performed on this to select only the biggest dark areas in the binary image. The binary mask and the final crypt

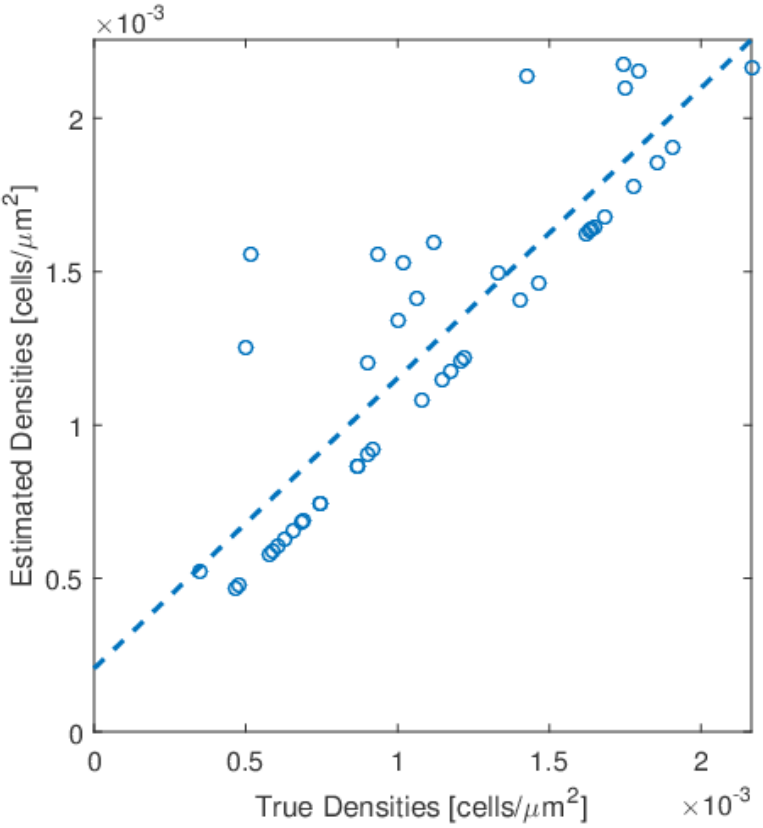

Figure 30: Estimated densities (y-axis) versus true densities (x-axis). The regression lines has $m = 0.8035$ and $q = 0.0001$, the value of R^2 is 0.761.

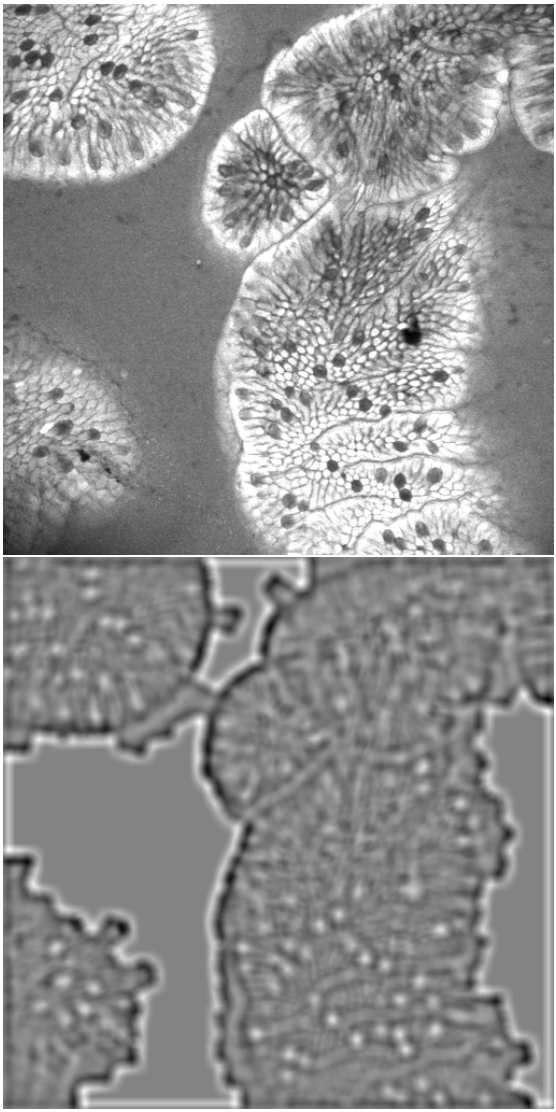

Figure 31: An image (top) and the response of the matched filtering among the image and a bi-dimensional filter, built to cause maximum response in correspondence of dark small regions of the given image.

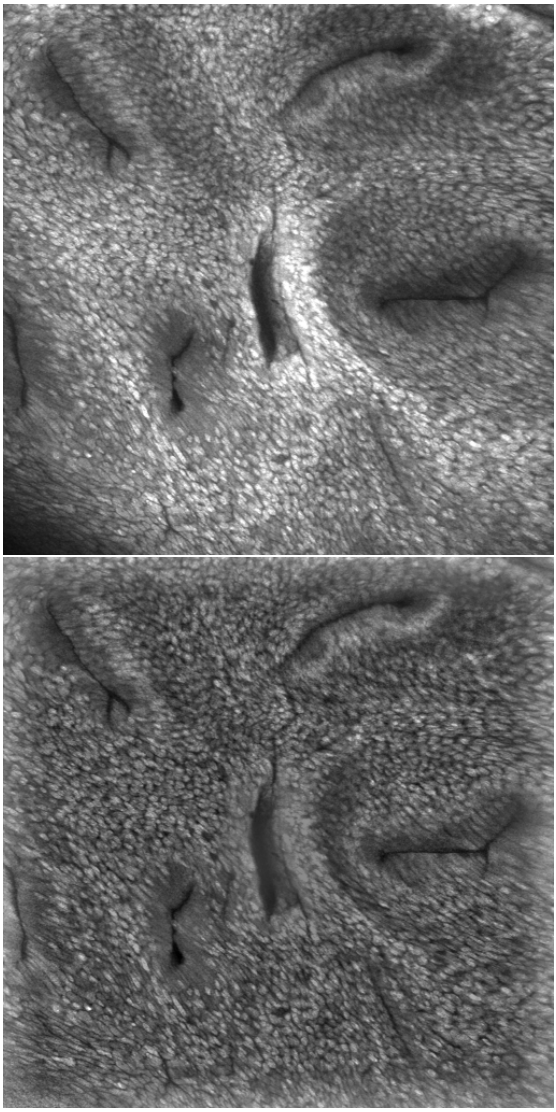

Figure 32: (top) An image exhibiting goblet cells. (bottom) A normalized version of the image.

segmentation of an image of the dataset is shown in Fig. 33.

3.6 Fluorescein leakage and cell dropout

As this chapter's introduction presented, defects in intestinal barrier function have been associated with various inflammatory intestinal diseases such as Crohn's disease, ulcerative colitis and irritable bowel syndrome. It is therefore essential that an intact intestinal barrier separates the host from the intestinal contents. It is estimated [75] that epithelial cell shredding in mice leaves a gap or discontinuity in the epithelium that in most cases resolves in 8-12 minutes, potentially challenging the epithelial barrier [90]. With acriflavine (or fluorescein) and CLE, these gaps can be detected in humans [67; 75]. Cell shedding is increased by high concentration of tumour necrosis factor (TNF), suggesting that this process might be increased in IBDs [91]. In fact, direct visualization of barrier defects is possible with CLE, and these parameter predict relapse of IBD [75]. Quantitative ways of measuring in real time gaps, micro-erosions and fluorescein leakage would improve the clinical practice. In the proposed CADSS, a hyper-fluorescence detector is built-in into the GUI. Along with all other quantitative measurements proposed and introduced in this Chapter, it highlights and contours the regions in each image that show almost white-saturated intensities (over 95% into the image domain scale space), provided that their areas are significantly big (more than 200 pixels). A simple thresholding is done on the input image, obtaining a binary mask highlighting the almost-saturated regions. On this, a simple morphological closing operation is performed to join close uniform leakage regions, and all regions with an area bigger than 200 pixels are kept. An example of this is shown in the images in Fig. 34. Along with all other methods implemented in the GUI, it will be presented in the next Chapter.

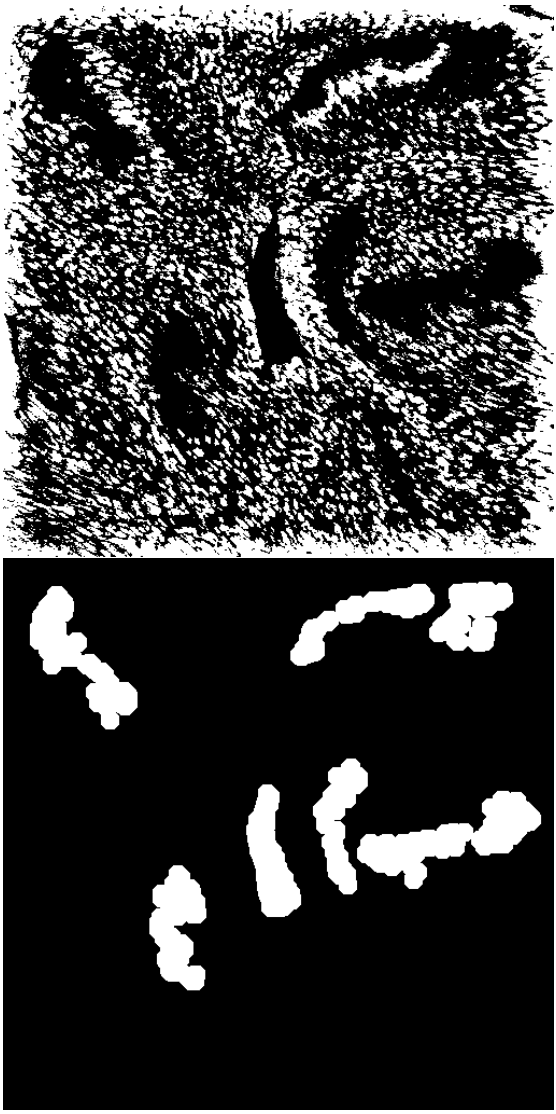

Figure 33: (top) unprocessed binary transformation of the edited image shown in the previous Figure. (bottom) The final binary segmentation mask.

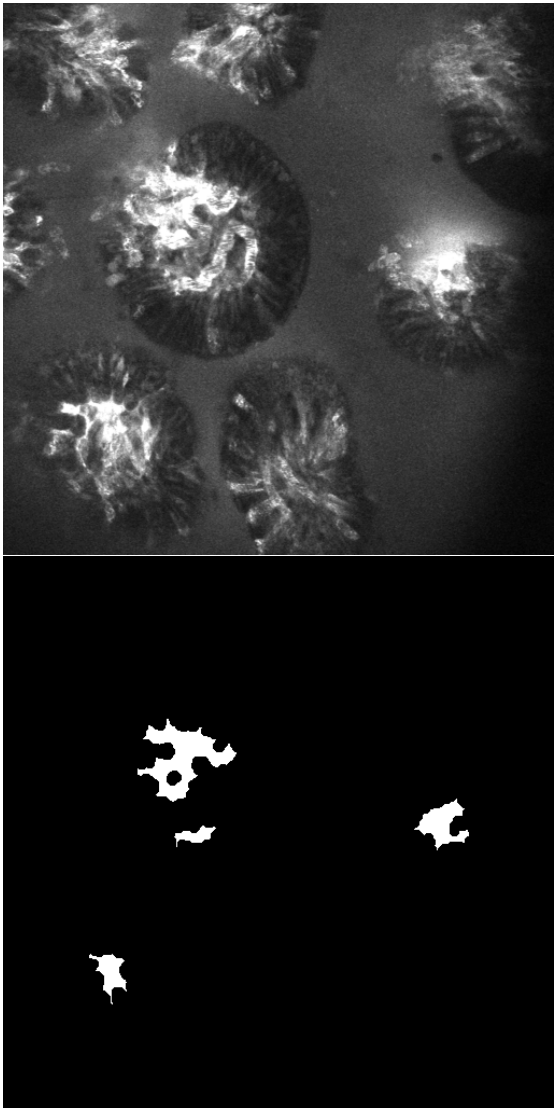

Figure 34: (top) An image with fluorescein leakage in the intervillous space. (bottom) Regions most affected by leakage according to the proposed detection scheme.

Chapter 4

Graphical User Interface

4.1 Overview and system requirements

To better showcase the results, and to be the most useful to medical experts, a graphical user interface (GUI) has been implemented in MATLAB, using GUIDE. The GUI starts by asking the user to load the image requiring analysis (Fig. 35), displaying then the main interface (Fig. 36).

The GUI does not need external toolboxes to work correctly: to run properly, it only requires an Internet connection to download the MATLAB runtime with all the necessary packages. An offline version is available, bundled with MATLAB runtime.
Along with the main executable, *CLE Inspector* comes with five demo images to showcase the software's capabilities. To keep the size of the package small, the multiscale analysis has been discarded, in favor of a faster (and more lightweight) single-scale analysis, based on the half-size scale of the original images (512×512). The GUI is itself split in two parts: on the left, the space is reserved for visualization purposes. Once loaded, the image is shown in this space. On the right, as it's visible in Fig. 36, the options to customize the desired quantitative measures to compute are presented to the user. The top part of the options panel highlights the input toggles needed before starting the analysis. The bottom part, instead, offers the user the tools to visually represent the computed metrics, and

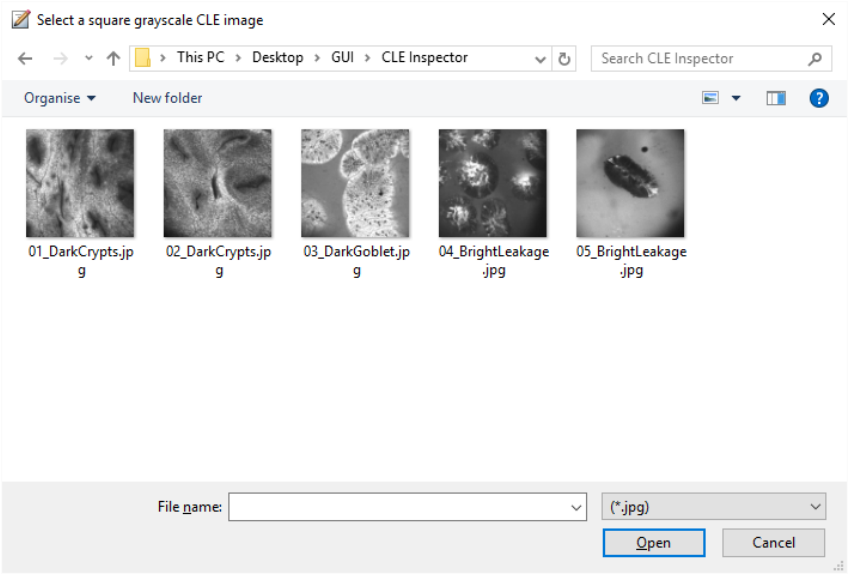

Figure 35: The user is asked to select an image to load.

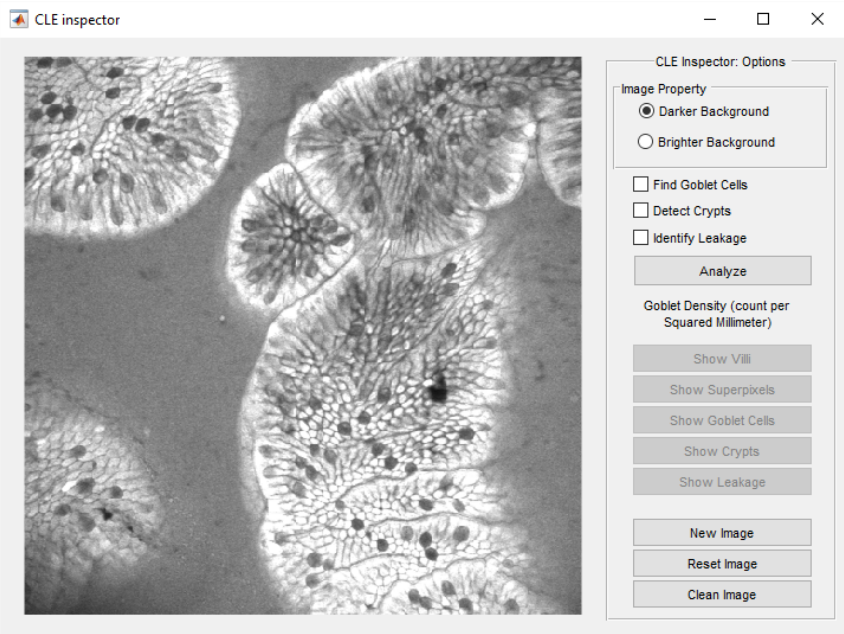

Figure 36: The main User interface of *CLE Inspector*.

present the (eventually) computed goblet cells density. In particular, all the features presented in Chapter 3 are computed (goblet cells density, leakage and hyperfluorescence, villi segmentation and eventual detected crypts). In the next sections inputs, outputs and the whole workflow of the process (along with the details) will be presented, with visual examples (screenshots of the GUI) concluding this Chapter.

4.2 Inputs

According to the very nature of the image under analysis, the user selects if such image exhibits a background that is darker or brighter with respect to the villi. This choice is required (on default, *Darker Background* is selected, since it's the class with the highest number of elements on the training set): according to this input, the algorithm selects the best classifier for the villi segmentation process. More details will be presented in the next paragraph. The other inputs required are the selection of the image features on which the computation should be based upon. Three different optional processes are offered to the user, that needs to select the checkboxes of the processes that he wants *CLE Inspector* to run. When all the input parameters have been set, the user clicks the *Analyze* button, and the computation starts.

4.3 Processing and workflow

When the *Analyze* button is pressed, the first step is the loading of the correct random forest for the learning-based villi detection step. In fact, two different random forests have been trained (on two different training sets) and included in *CLE Inspector* for the purpose of generalization. The two different datasets were build by labeling each image to a darker or brighter background class according to the difference between the mean intensity in the villi and outside of the villi, based on the manual segmentation binary mask given by medical experts. A brief presentation of the two classifiers is shown in Table 2.

	Darker RF	Brighter RF
#Trees	50	50
#Trained Superpixels	77262	54125
#Trained Images	181	127
#Superpixels - Villi	63880	34947
% of SP-Villi	82.7	64.6

Table 2: Information about the training process of the two classifiers that are bundled in *CLE Inspector*.

Three different measures have been computed for analyzing the different random forests. The first two are dependent on the number of trees included in each random forest, and the last one is reflective of the features chosen. Both of the first two measures are an indication of the robustness of the training process: these are the Out-of-bag error and the mean margin. The OOB error is a measure of the misclassification probability for the out-of-bag observations in the training data. The plots in Figs. 37 and 38 show the cumulative error obtained while predicting the class of the out-of-bag observations using only the number of trees shown in the horizontal axis. Apart from an unexpected spike corresponding to the first trees in the OOB Error graph for the darker background RF, both graphs show as expected a decreasing trend for the error as long as more trees are included in the classification process. Moreover, there is a difference in the performance of the algorithms: the final average error for the brighter-background RF is 13.03%, while the one for the darker-background RF performs better, at 10.52%.

The same comparison has been done to each classifier's mean margin. Margin, in the binary classification field, is a measure of the distance between the predicted probability for the positive class and the predicted probability of the negative class. Obviously, the higher the margin, the better the classification results. As shown in Figs. 39 and 40, the conclusion from the previous analysis is confirmed. In the bright background case the random forest reaches a stable margin (of about 0.553) when only 10 to 15 trees are included in the classification process.

In the darker background case the situation is similar, but the mar-

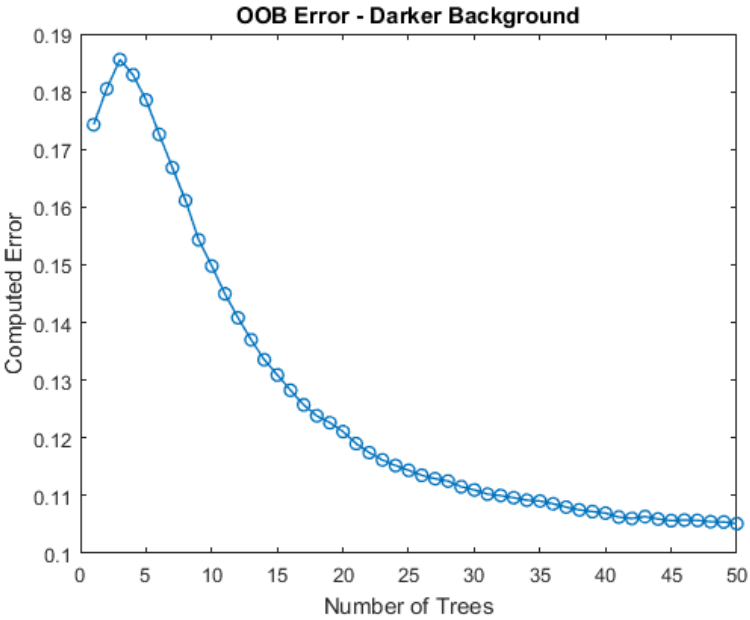

Figure 37: How error in the out-of-bag observations evolves, as more trees are included in the random forest ensemble for the darker background images.

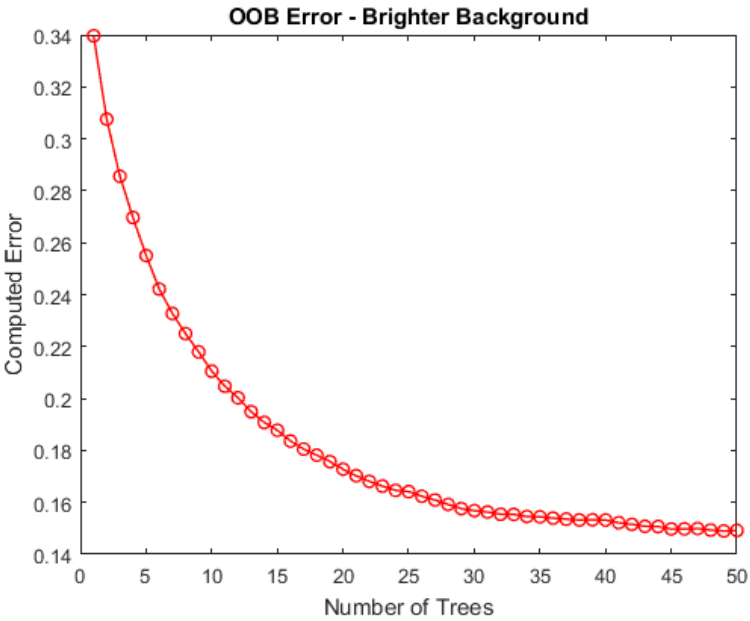

Figure 38: How error in the out-of-bag observations evolves, as more trees are included in the random forest ensemble for the bright background images.

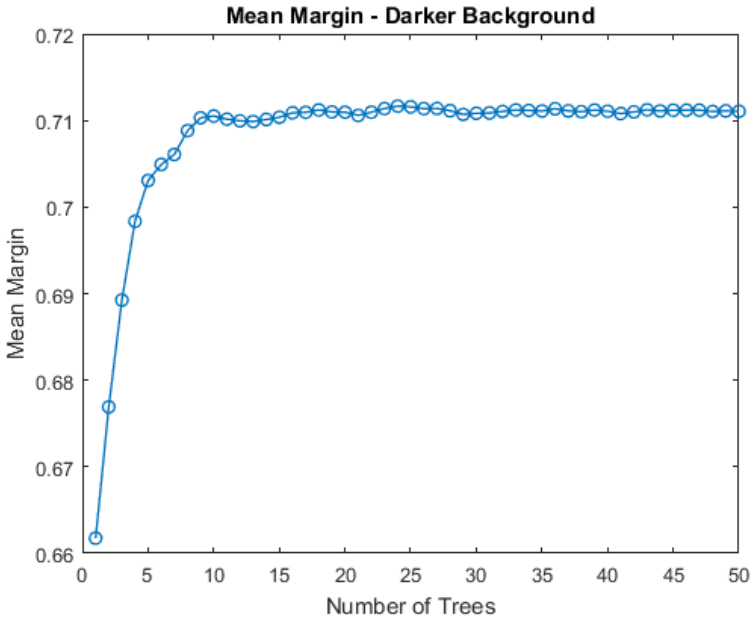

Figure 39: How the mean margin in the out-of-bag observations evolves, as more trees are included in the random forest ensemble for the darker background images.

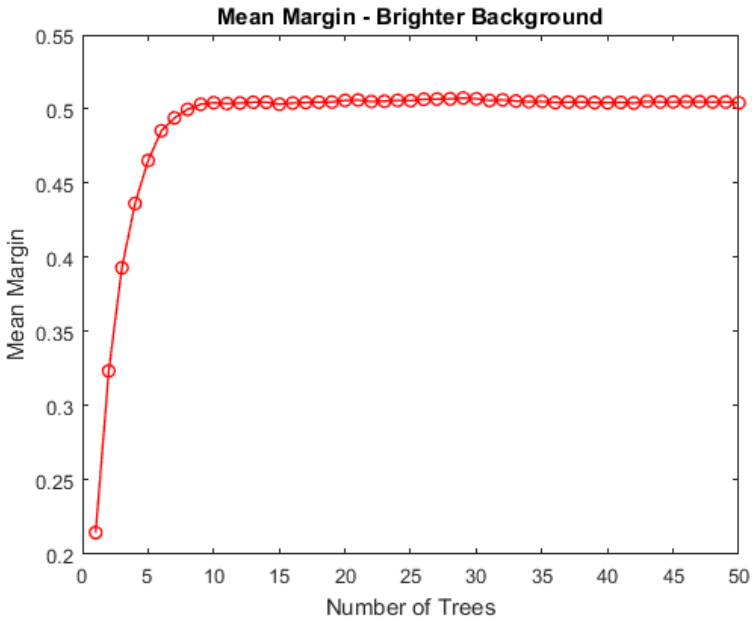

Figure 40: How the mean margin in the out-of-bag observations evolves, as more trees are included in the random forest ensemble for the bright background images.

gin is higher (at about 0.648), as it is expected given the previous point made about the OOB error. Given the combined information from these analyses, all 50 trees were kept for both the random forests.

Once one classifier is loaded, the input image is processed for features extraction. First, superpixels are computed from the image. For this step, the already introduced *vlfeat* has been discarded in favor of a new native MATLAB's image processing toolbox method to compute superpixels, introduced in the latest release (R2016a). This method comes with three positive points: it permits faster computation times; the GUI has no external dependencies; SLIC can be updated to *SLIC0*. With *SLIC0*, in fact, the only parameter to set is the desired number of superpixels to compute (that has been set to 441 for continuity reasons), a number that is not guaranteed to be the exact number of different superpixels computed by the end of the process, since the algorithm self-adapts to the image under analysis. The algorithm will vary the size of each superpixel, not forcing it to be close to a given value, using more degrees of freedom if compared to the simple SLIC version implemented by *vlfeat*. Because of all these reasons, this new native algorithm has been adopted in this GUI. For each of the 441 (on average) superpixels of each image, 37 features are extracted: 32 of those are the bins of the histogram done on the Local Binary Patterns extracted from the superpixel, using an adapted version of *efficientLBP*. Two other features are mean intensity and standard deviation of intensities in the superpixel. The last three features are contrast, homogeneity and energy, that have been computed via grey-level co-occurrence matrix statistical analysis. All these features are well known and widely used in the computer vision field, have already been presented in the previous Chapter and convey information about the texture of the region in which they are computed. To study and rank all 37 features, the so called Permuted Predictor Delta Error has been computed as a third measure for analyzing the correctness of the selected random forest. This measure expresses each feature's importance. For each of the 37 variables, the measure is the increase in the prediction error if the values of that features are permuted, in all the out-of-bag observations. Since a random forest is an ensemble of decision trees, the measure

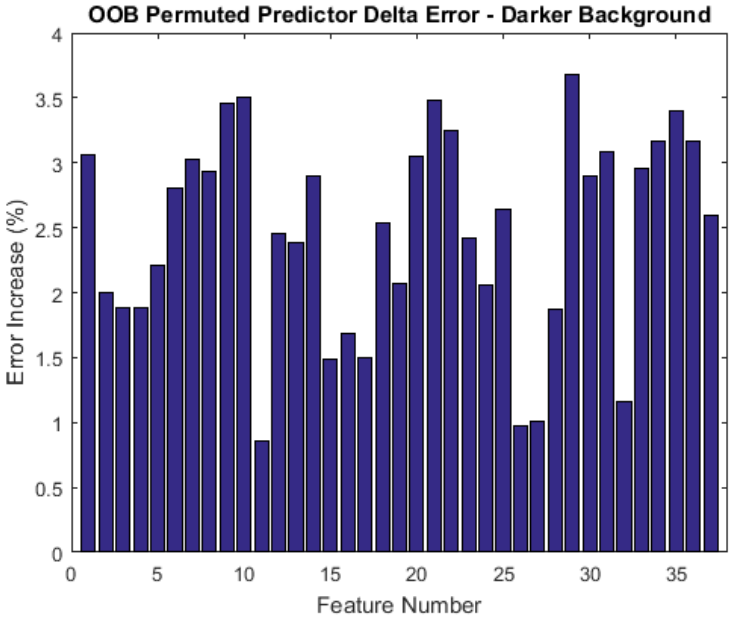

Figure 41: A bar plot highlighting each feature's importance in the darker-background RF bundled in *CLE Inspector*.

is computed for each of the 50 trees, then averaged on the entire ensemble and divided by the standard deviation over the entire forest. The two different rank plots are shown respectively in Fig. 41 and Fig. 42. For the random forest used for brighter-background villi segmentation, it is clear that the most important feature is the first one, that corresponds to each superpixel's mean intensity. Although some bins are clearly less informative than others, each feature if permuted increases the prediction error, as expected. For the second random forest, the one trained on the darker background images, the results are different. The mean intensity of the superpixel, in this situation, is not among the first 5 most informative features. This time, the most important features are among some of the bins from the LBP histogram. From this, it can be inferred that for images in which the two classes possess both mean intensities (darker

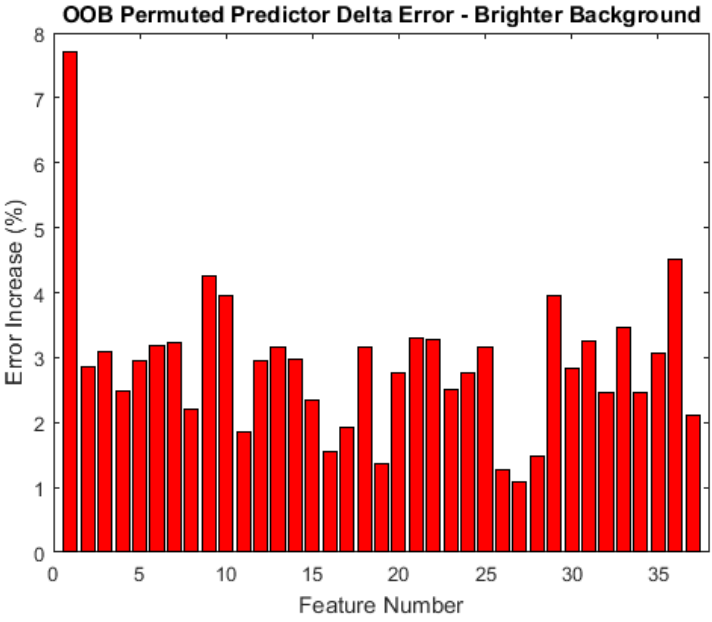

Figure 42: A bar plot highlighting each feature's importance in the brighter-background RF bundled in *CLE Inspector*.

villi, brighter background) and texture (more textured villus, smoother background) that are different enough, both classifiers should be able to distinguish well enough villi from the background. An example of this will be shown in a comparison in the demo examples.

After feature extraction, the labeling is performed on the input image by feeding the extracted feature matrix to the selected classifier (one row for each superpixel), that predicts if each of them is part of the villous fold or not. Given the prediction, a superpixel-based binary mask for villi detection is reconstructed, labeling the villi. A post-processing step is computed, that keeps in the mask only the connected regions with a pixel area of at least 65×65 pixels (4225-pixel area, correspondent to the hard threshold of 130×130 pixels put on the method presented in Chapter 3 that was dealing with images 4 times bigger (1024×1024 w.r.t. 512×512 in this case), to exclude isolated false positives.

Then, if the user checked the *Find Goblet Cells* checkbox, a goblet cells detection algorithm is run, to calculate both the goblet cells density and a binary mask for visualization purposes. First, the image is normalized (zero mean, unitary variance). Then, a filter for matched filtering is devised, such that the filter response is maximal in correspondence to small dark round structures (the filter size has been set to 11×11 pixels). After this, the matched filtering is performed by convolving the input image with the filter. This response is generalizable, and always conveys the same information: in fact, the regions in which small dark structures are located in the input image corresponds to the regional maxima in the filter response. Because of this, a simple hard threshold process, coupled with a filtering on area size, gives us a reasonably accurate estimation of the location of the goblet cells in the input image. Obviously, this information is coupled with the villi detection results, so that goblet cells are only selected in the area that the random forest has labeled as villous mucosa. The only downside of this sub-process is due to the image normalization, that inserts artifacts close to the image boundaries, therefore sacrificing a bit the accuracy of the detection in those regions.

If selected, then, a crypt detection algorithm is applied. This method is based on simple thresholding on a normalized version (as above, based

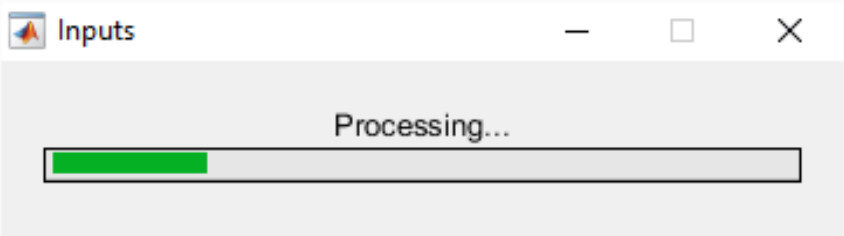

Figure 43: A waitbar informs the user about the state of the processing step.

on local standard deviation filtering) of the input image. The image, this time, is normalized according to the minimum and the maximum values of the regions of the image that have been selected as parts of the villous folds. By doing so, the darkest areas of the image (supposedly, crypts and goblet cells) are easily highlightable with a simple binarization process. Then, a filtering is done on the binary mask to keep only areas larger than 1024 pixels (32×32 patches).

The process concludes with a leakage detection mechanism, that simply creates a binary mask that highlights the regions that are saturated in the input image (over 95% of the grayscale permitted range in the non-normalized input image).

During all of the steps above, the user is given a waitbar (that is updated linearly with time as the analysis is performed) to convey a sense of the status of the processing. This is shown in Fig. 43. Its update is linear with the feature extraction process (the update step size corresponds to a computed superpixel), since this is the computational bottleneck of the whole procedure.

All the steps in the pipeline, from the press of the *Analyze* button to the graphical refresh of *CLE Inspector* main window after the analysis, are computed in a relatively fast time, as shown by Table 3 (average times computed on the 5 demo images in 5 different runs). The laptop that has been used for calculating computation times is a Intel i7-3630QM@2.40GHz CPU, 8 GB DDR3 RAM, Windows 10 Pro OS with a Samsung 840 EVO SSD of 1TB.

	Time (s)
Random Forest Loading	0.84
Superpixel computation	1.33
Feature extraction	14.22
Prediction + Villous Mask	0.44
Goblet Cells	0.28
Crypts	0.26
Leakage	0.02
Total (5-runs average)	17.39

Table 3: Computation times for 5 average runs on all demo images in a medium-tier laptop.

4.4 Output

When the processes described above conclude, *CLE Inspector* outputs one textual result and, according to the user's choices, several user-friendly visual options. The textual result is the goblet cell density: it is estimated by counting the goblet cells in the field of view. Quantitative measures are very dependent to the image under analysis: images with heterogeneous structures in their field of view can bring to erroneous calculations, and report therefore biased results. On top of this, for this calculation it has been supposed that the image under analysis is an image with in-plane resolution of 0.5×0.5 mm and dimensions 1024×1024 pixels. If the *Find Goblet Cells* checkbox has not been selected, a message (*Goblet Density: Not computed*) is shown to the user.

Visual results, instead, are managed through a series of push buttons, that possess these properties:

- *Show Villi*: as the name says, this button creates a green transparency overlay of the villi binary mask on top of the original image. This item is always clickable, once the process has finished computing.

- *Show Superpixels*: a new overlay is done on the input image, showing the borders of the superpixel's analysis. Again, this item is always clickable, too.

- *Show Goblet Cells*: if the user did check *Find Goblet Cells*, this button will be enabled, and will show an overlay of the binary goblet cells segmentation mask in orange on the original image.
- *Show Crypts*: this button is available only if *Detect Crypts* has been checked before clicking the *Analyze* button. This overlays a cyan transparency layer on top of the original image showing the estimated crypts.
- *Show Leakage*: again, the button is clickable only if *Identify Leakage* was checked, and by clicking it a red overlay on the saturated regions is shown on top of the original image.
- *Clean Image*: a fast way to restore the original image in the figure space, that lets the user start a new series of overlays using the already-performed computations if an undesired operation was performed inadvertently.
- *Reset Image*: this button lets the user clean the image and choose the parameters to perform a new analysis of the image already loaded.
- *New Image*: pushing this button will open a windows prompting the user to choose a new image to be loaded.

4.5 Demo examples

The software comes with five demo images, anonymized and reflective of the heterogeneous nature that a CLE-aimed algorithm needs to be able to treat (and possibly exploit). In the rest of this chapter, figures illustrating the performances on all five demo images and the GUI itself, its features, characteristics and behavior will be shown.

The first figures (Figs. 44, 45, 46, 47, 48) show how the UI behaves, according to the checkboxes the user selected before the analysis began. In this case, since the user selected *Darker Background* and *Find Goblet Cells*, only the *Show Goblet Cells* button is clickable (along with *Show Villi*, *Show Superpixels* and *Clean/Reset/New Image*, that are always clickable).

Figure 44: The main results interface: the computed goblet cells density is given to the user, and the different buttons give the user the chance to visually present the results.

The output can be seen in the text box between the *Analyze* and the *Show Villi* buttons, and the buttons can be clicked accordingly to combine the results, as shown in the images. Pushing a button adding an overlay on top of the image will hide previously drawn features: clicking *Show Goblet Cells* and then clicking *Show Villi*, for example, will hide the goblet cells. It's enough to click *Show Goblet Cells* again to overlay the goblet binary mask on top of the villi.

The second batch of images show how the algorithm behave on two different images from a patient suffering from Crypt Hyperplasia. The user instructs the software to detect crypts (Fig. 49), and after the processing it can be seen that the text output regarding goblet cells (Fig. 50) outputs a "*Not Computed*" message, and the buttons updates (enabling

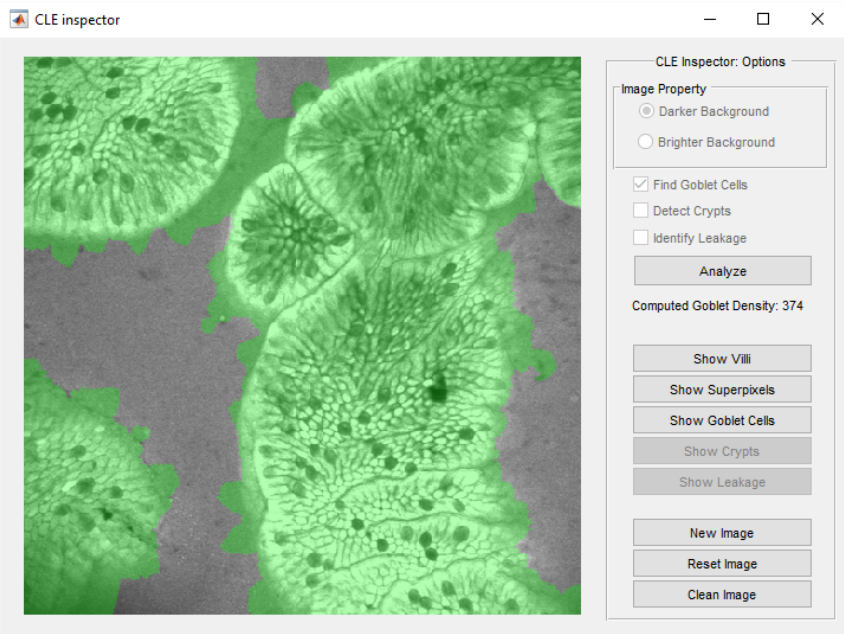

Figure 45: The result of the villous segmentation.

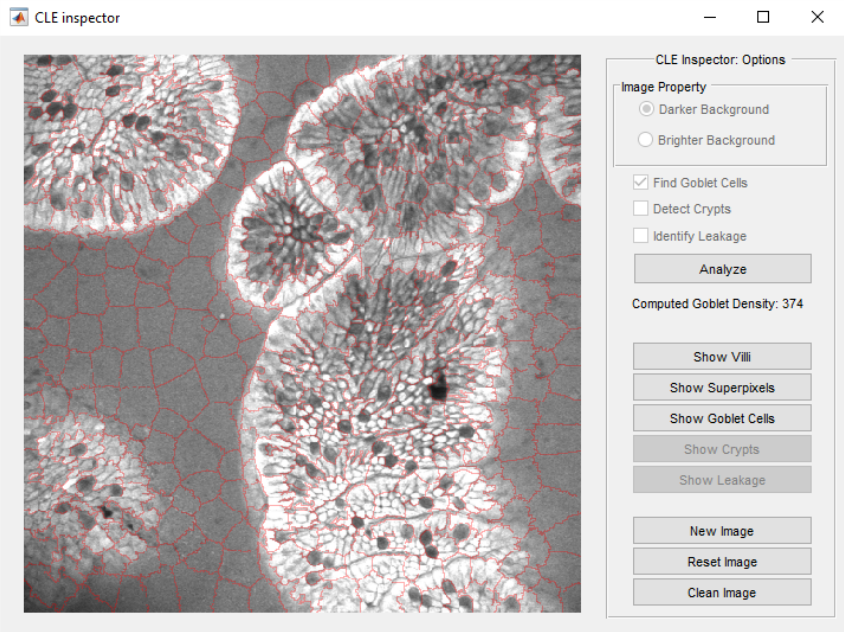

Figure 46: Computed superpixels of the input image.

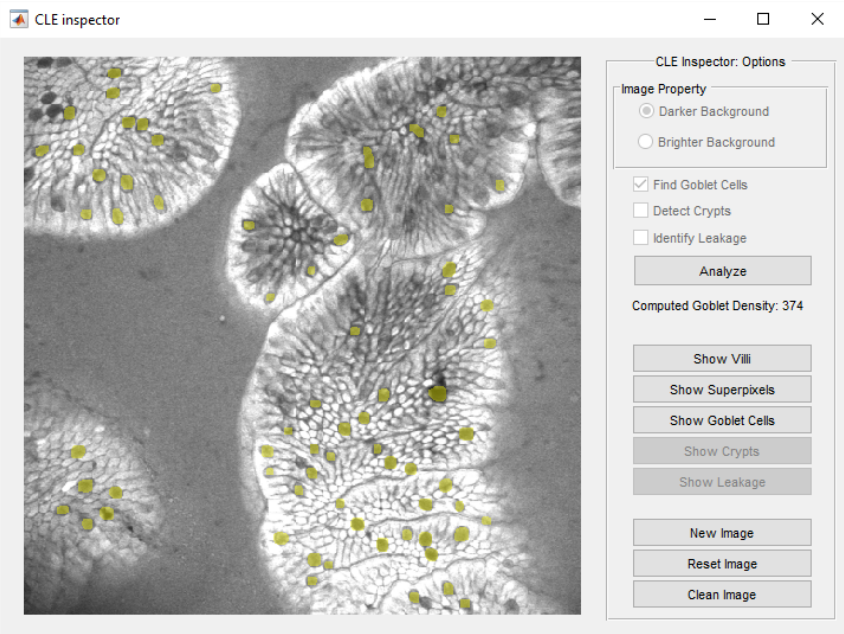

Figure 47: Goblet cells detection, based on the fast matched filtering algorithm.

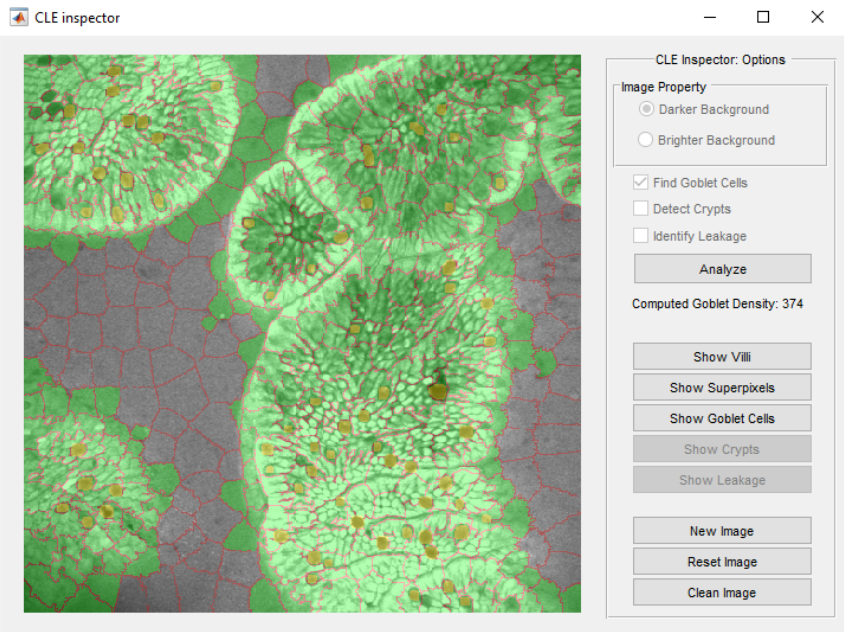

Figure 48: All the results combined from this image.

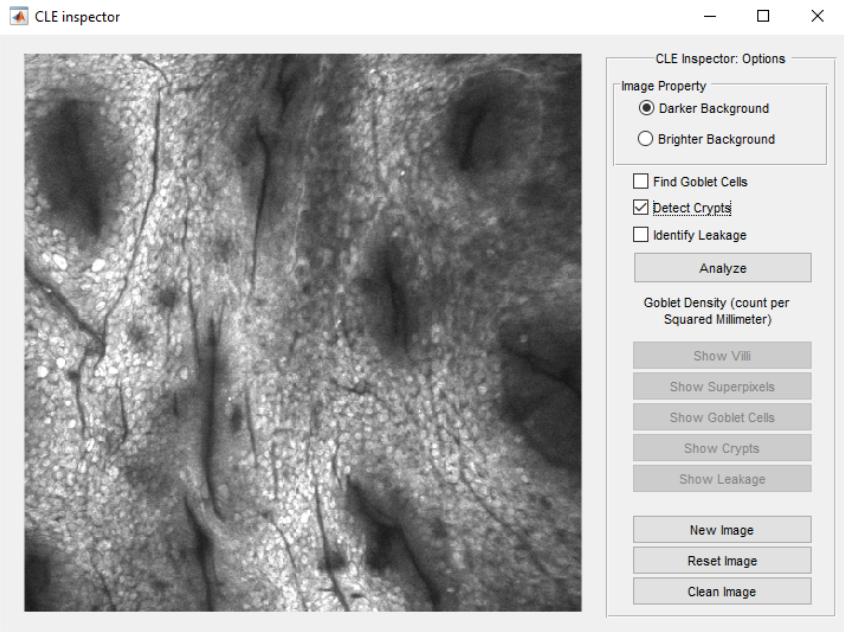

Figure 49: The main User interface of *CLE Inspector*, with an image exhibiting crypts.

and disabling themselves) accordingly. Again, the user can combine the visual outputs as he wishes (Figs. 50, 51, 52, 53, 54, 55).

When leakage detection is needed, instead, the user needs to select *Identify Leakage* in the corrispective checkbox, after having chosen the correct property of the image (darker or brighter background). Again, the buttons update accordingly (Figs. 56, 57). In these situations in which the background is fairly dark and smooth with highly textured villi, the two random forests behave similarly: this can be seen in Figs. 58 and 59.

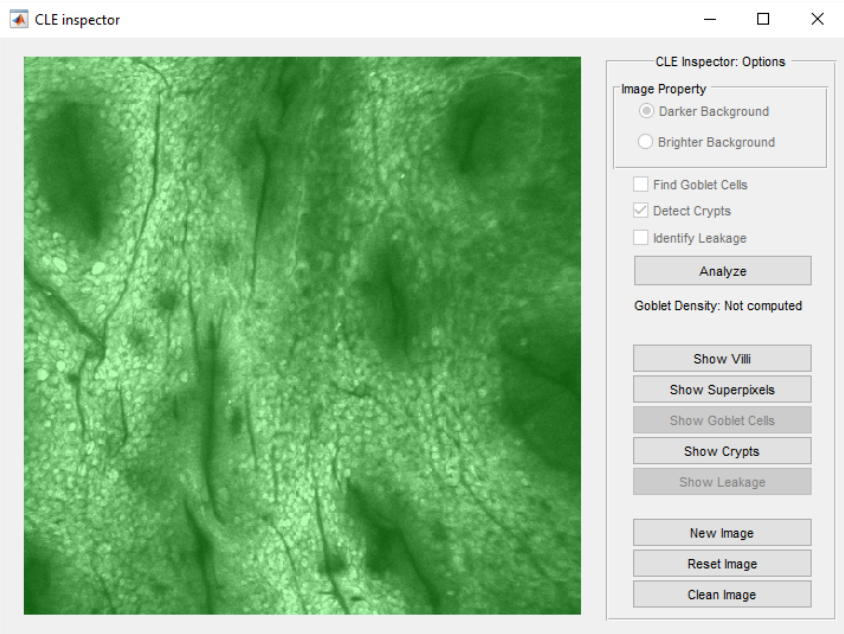

Figure 50: Villi segmentation.

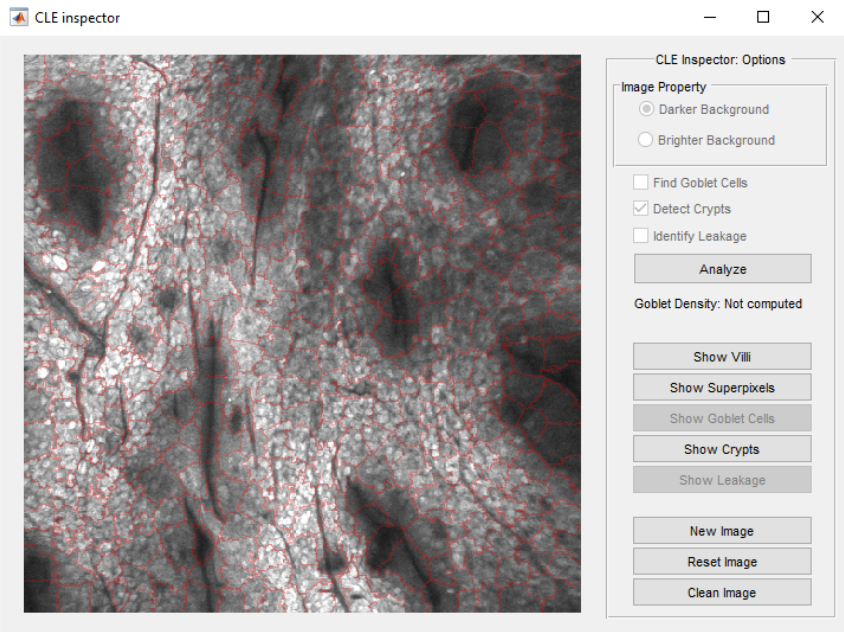

Figure 51: Computed superpixels.

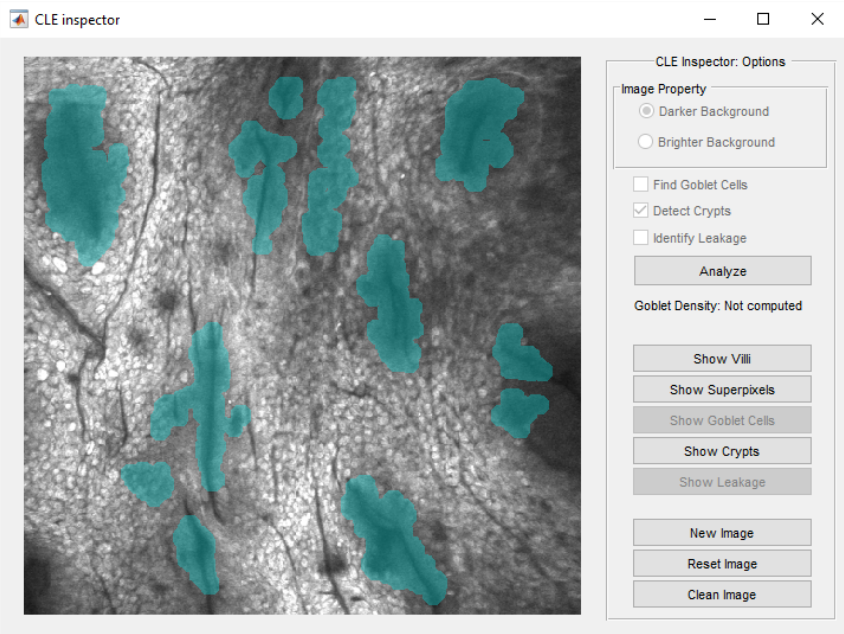

Figure 52: Detected crypts.

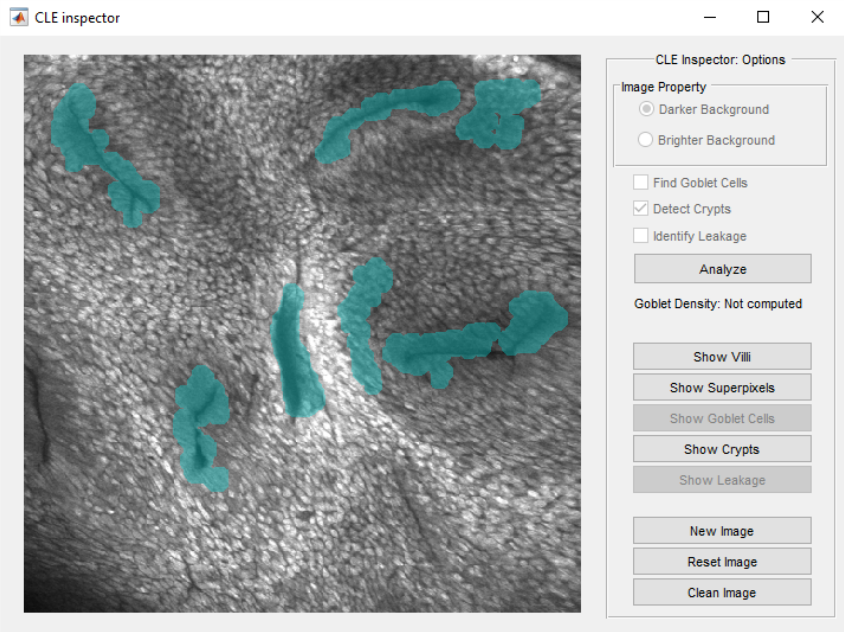

Figure 53: Another image showing crypt hypertrophy.

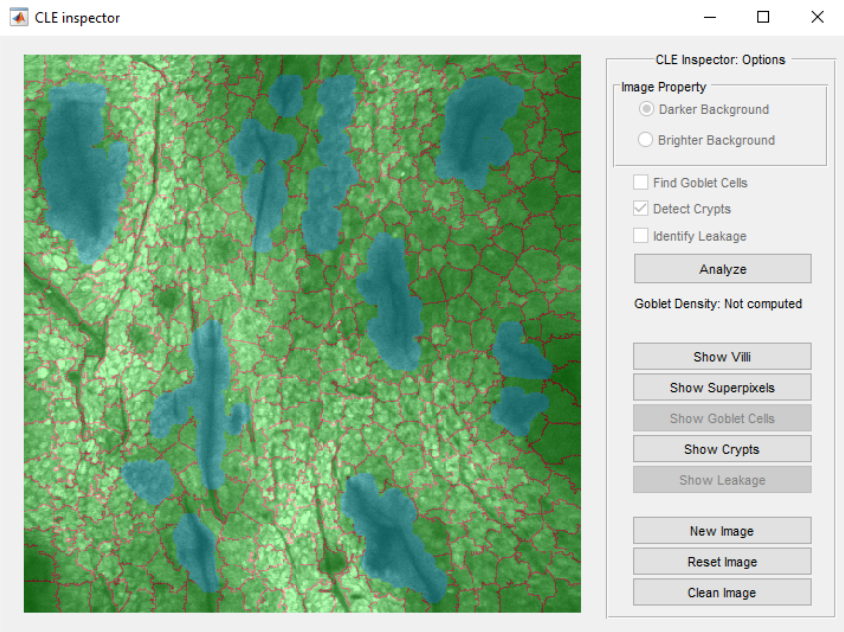

Figure 54: Cumulative results from this never before seen image.

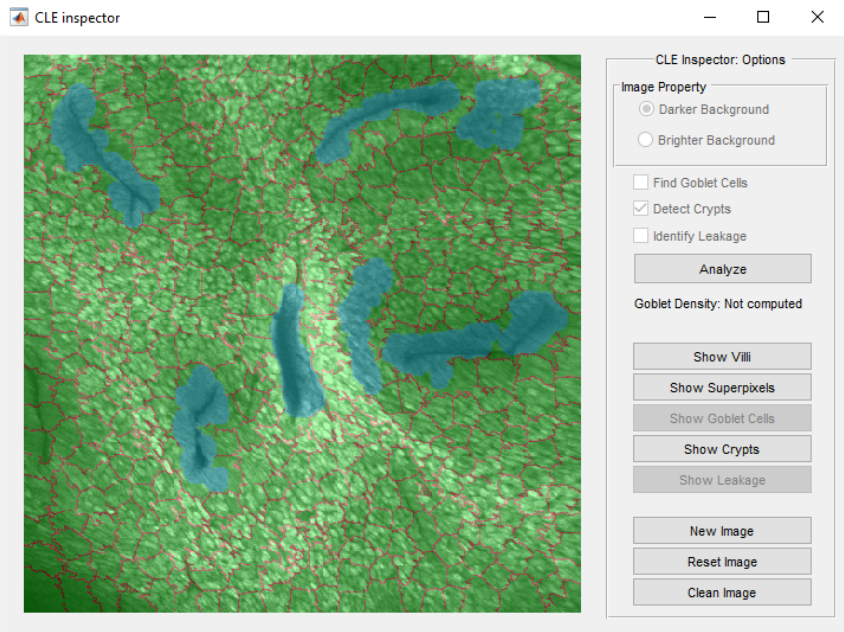

Figure 55: Cumulative computed results.

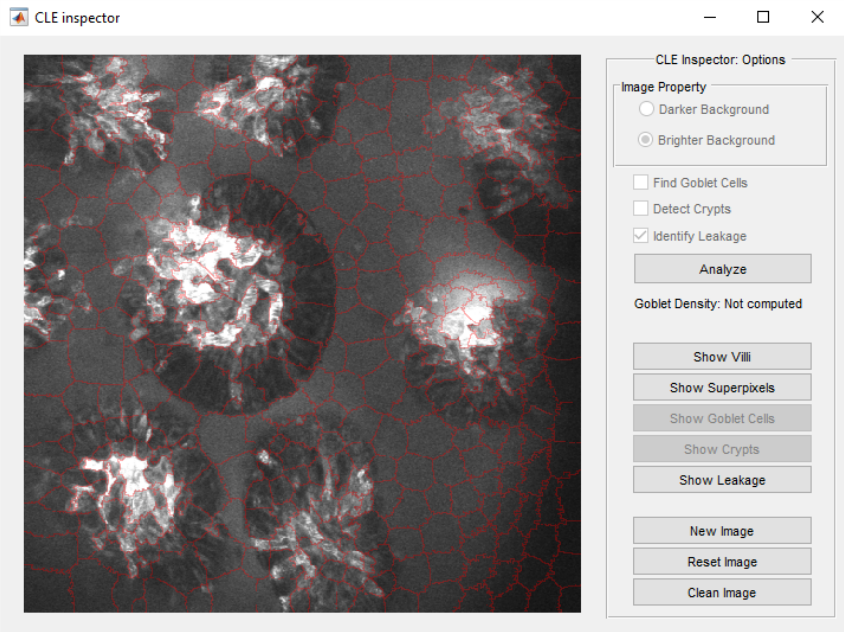

Figure 56: Computed superpixels.

Figure 57: Regions with hyperfluorescence have been detected by *CLE Inspector*.

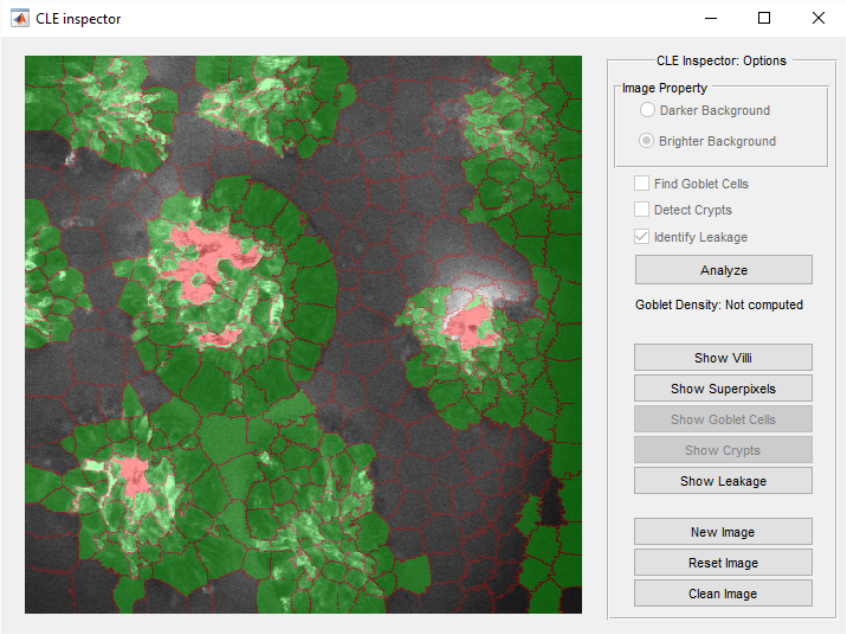

Figure 58: Cumulative results on this image using the leak-based random forest.

Again, a comparison among the two random forests on another image showing smooth background against textured (albeit darker in most superpixels) villus is shown in Figs. 60 and 61.

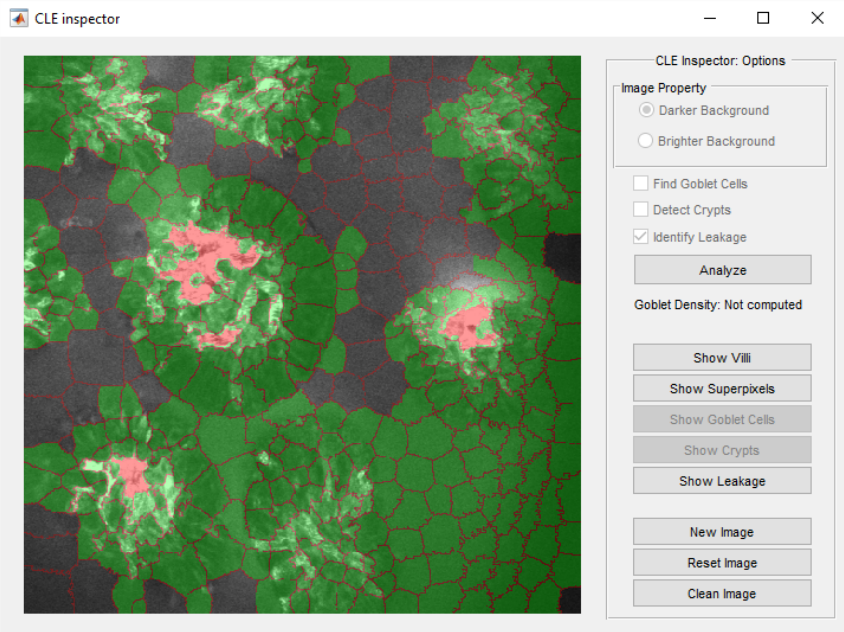

Figure 59: Results on the same image, using the darker background-based random forest. The results are worse, but comparable.

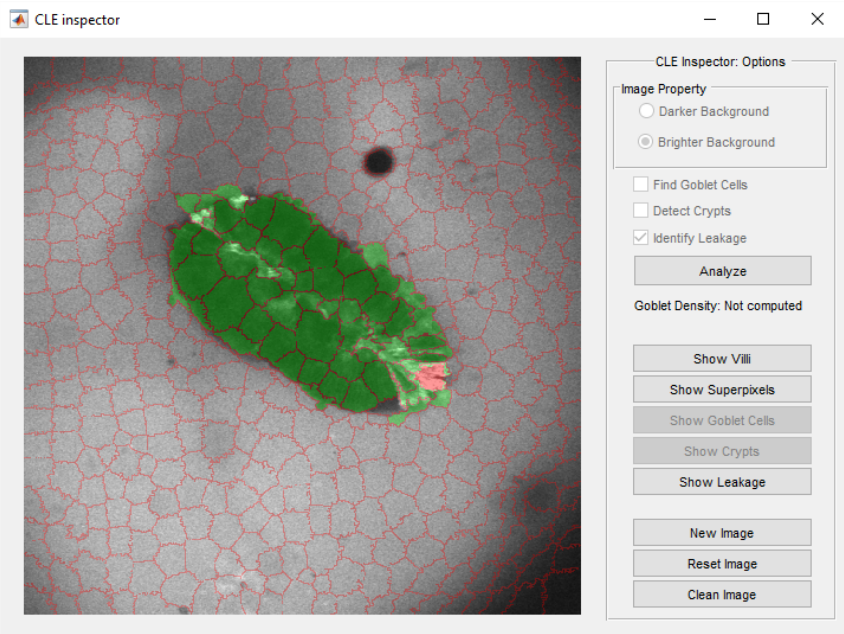

Figure 60: All results computed using the leak-based RF.

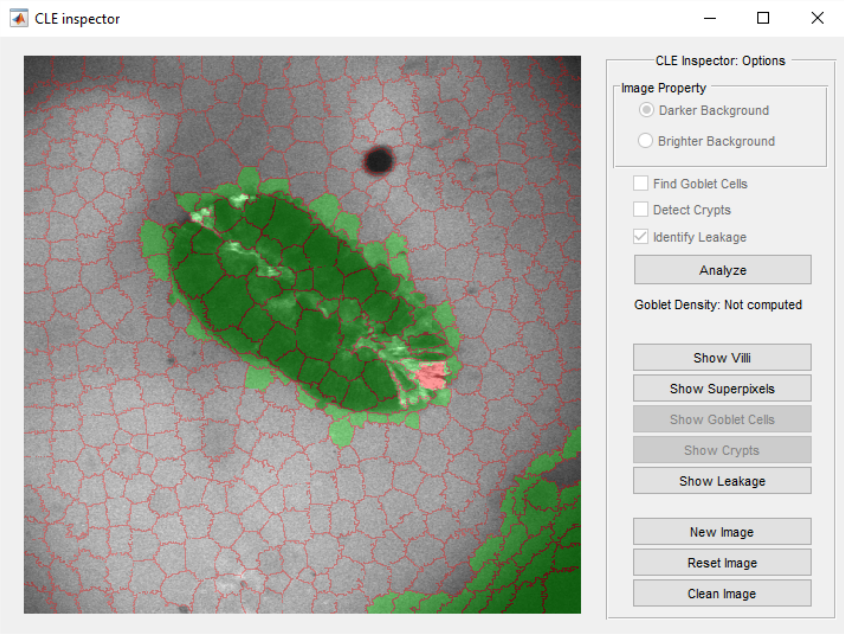

Figure 61: All results computed using the darker background-based RF. A false positive has been included in the villi segmentation.

Chapter 5

Conclusions

In this thesis, the first methods for quantitative analysis and salient features detection for confocal laser endomicroscopy images have been presented. All methods have been presented singularly in Chapter 3, while a comprehensive software (*CLE Inspector*) has been developed and presented in Chapter 4. All these methods have the aim of solving a difficult problem, that is salient features detection in medical images originating from CLE. In Chapter 2, the state of the art of this field has been reviewed: the state of the art is improvable by a lot, since no quantitative methods have been proposed to solve the problems affecting the medical experts in the field. Extreme heterogeneity of structures, alterations and reactions to contrast agents all have impeded the successful application of image processing techniques in CLE images. It has been explained why a future has been envisioned in which, thanks to quantitative analysis of CLE images, virtual biopsies will be possible without the need to recur to histopathology. In Chapter 3, all the algorithms envisioned, implemented and tested during this doctorate have been proposed. Two different algorithms for detecting villi in CLE images have been presented: the first relies on morphological processing and suffers from unforeseen heterogeneity of the images, while the second relies on learning and the application of SLIC superpixel segmentation, a computer vision technique based on K-means clustering used to compute a

rough first-step segmentation of the images in multiple uniform regions to classify according to texture-based features computed on the pixels belonging to them. Two algorithms for detecting goblet cells in the villous folds have been presented: an algorithm based on matched filtering and Voronoi's segmentation, and a faster one only based on matched filtering and hard thresholding, that offers results comparable (if not better) with the first one. On top of this, a crypt detection algorithm has been implemented for images from patients suffering from crypt hyperplasia (one of the main alterations of celiac disease, among the others), and a leakage detection for images originating from patients with altered intestinal permeability. In Chapter 4 a software has been proposed that implements the methods presented in the previous chapter, that requires an image in input and computes quantitative analysis of such image. The user can visually see the results of all the different part of the algorithm overlaid on the original image, as long as a quantitative measure (if required and toggled) of the goblet cells density in the input image. The algorithms implemented in *CLE Inspector* are faster versions of the ones shown in Chapter 3, with reactivity and speed in mind. If the user has a multi-core system, the process takes advantage of that to speed up the process, too. In the GUI, in fact, a smaller number of features is used for villi detection (only one scale of the image is used, instead of the three presented in Chapter 3), and the main problem of heterogeneity among images has been tackled by letting the user choosing which features to compute, according to the nature of the image under analysis. On top of this, two different random forests have been trained to classify images showing either darker or brighter background with respect to villi (i.e., if leakage is occurring outside the villi's mucosa or not). In such a way, a more focused detection is possible, due to the a priori information given by the user.

The GUI (*CLE Inspector*) could be improved in various ways in the future (lose the need to download MATLAB's runtime to work, nicer GUI with C/C++ or Java, better trained random forests). A flag conveying information about the certainty of the output could be implemented, to tell the user if the input image differs too much from the images used for

training the classifier. In such case, in fact, the predictions could not be sound or informative enough, and false positives / false negatives might affect the accuracy of the model in unforeseen ways. Additional algorithms can be devised and included effortlessly in the framework (lamina analysis, different features on which to base the detection processes, different learning algorithms and so on), given its modular nature and programming easiness. A total reimplementation of this MATLAB GUI as a front-end web-app could be done in the future, to lose the dependency to MATLAB (or its runtime) for a successful run. For this, the security and anonimity of the images under analysis will need to be taken into serious account too, given that personal information and sensitive data need to be treated with precautions in the digital age.

Appendix A

CADSS for the detection of precancerous lesions in NBI endoscopy

A.1 Summary

Barrett's esophagus (BE) is a precancerous complication of gastroesophageal reflux disease in which normal stratified squamous epithelium lining the esophagus is replaced by intestinal metaplastic columnar epithelium. Repeated endoscopies and multiple biopsies are often necessary to establish the presence of intestinal metaplasia. Narrow Band Imaging (NBI) is an imaging technique commonly used with endoscopies that enhances the contrast of vascular pattern on the mucosa. A computer-based method for the automatic normal/metaplastic classification of endoscopic NBI images is presented in this Chapter. Superpixel segmentation is used to identify and cluster pixels belonging to uniform regions. From each uniform clustered region of pixels, eight features maximizing differences among normal and metaplastic epithelium are extracted for the classification step. For each superpixel, the three mean intensities of each color channel are firstly selected as features. Three added features are the mean intensities for each superpixel after separately applying to the

red-channel image three different morphological filters (top-hat filtering, entropy filtering and range filtering). The last two features require the computation of the Grey-Level Co-Occurrence Matrix (GLCM), and are reflective of the contrast and the homogeneity of each superpixel. The classification step is performed using an ensemble of 50 classification trees, with a 10-fold cross-validation scheme by training the classifier at each step on a random 70% of the images and testing on the remaining 30% of the dataset. Sensitivity and Specificity are respectively of 79.2% and 87.3%, with an overall accuracy of 83.9%.

A.2 Introduction

Gastroesophageal tumor is one of the most common cause of death among industrialized countries, with a high mortality rate due to the difficulties of early discovery and treatment. The main precancerous lesion is known as Barrett's Esophagus [92], a condition in which any extent of metaplastic columnar epithelium that predisposes to cancer development replaces the stratified squamous epithelium that normally lines the distal esophagus above the gastroesophageal junction (GEJ). In the last decades, there have been various controversies and debates about the pathogenesis, the management and the definition of this common disorder, and a clear consensus is still missing [93; 94; 95; 96; 97; 98]. BE develops as a consequence of gastroesophageal reflux disease, which damages the distal esophagus epithelium and stimulates healing through columnar metaplasia rather than through the regeneration of more squamous cells [99]. Endoscopically, the columnar epithelium and the squamous epithelium are distinguishable: the former has a pink color with a coarse texture, while the latter exhibits pale colour and glossy appearance. An example of this difference can be seen in Fig. 62. Repeated endoscopies and multiple biopsies are often necessary to establish the presence of intestinal metaplasia, since the visual confirmation of the presence of columnar epithelium above the GEJ is not enough to diagnose BE in a patient. Generally, in addition to that, a histological confirmation of columnar metaplasia in esophageal biopsy specimens is necessary, although there

is discussion about which kind of metaplasia should lead to the final BE diagnosis (namely, if goblet cells are required or not to diagnose BE with certainty) [100].

Various imaging techniques can be used to screen the status of the gastrointestinal tract, but the most used are conventional white light endoscopy (WLE) and narrow band imaging (NBI). WLE uses a conventional RGB filter, a light source and a sensor to record high resolution images. Narrow Band Imaging, instead, uses an additional filter that splits white light into two specific lights with narrowed bandwidths (blue and green, at 400-430nm and 530-550nm respectively), canceling the contribution of the red light. In this way, since blue and green lights possess more superficial penetration than the red light, pit patterns and vasculature texture are enhanced in the resulting images. Automatic classification of endoscopic images can be used by experts as a support system, narrowing down the critical regions in the images, where abnormalities are most likely to be discovered, giving the experts a chance to perform targeted biopsies, instead of performing potentially damaging and not targeted biopsies using the four-quadrant protocol. This work presents a novel local feature extraction method, by first performing a rough clustering-based segmentation using superpixels, and classifying each superpixel according to the calculated features in each of them using random forests, a well-established classification method. The main aim of the method is the classification of each superpixel as normal or metaplastic, to be used as first step in a CADSS with the purpose of helping experts grading the severity of the patients.

A.3 Materials

In this study, 116 NBI images were obtained from clinical checkups conducted at Istituto Oncologico Veneto (IOV) in Padova, Italy, in which each patient underwent a surveillance endoscopy (Olympus CV-180). Each image's resolution is 720×480 pixels. In order to provide a ground truth, all images have been manually analyzed, providing an outline of the eventual lesion in each of the images. Typical images from the dataset

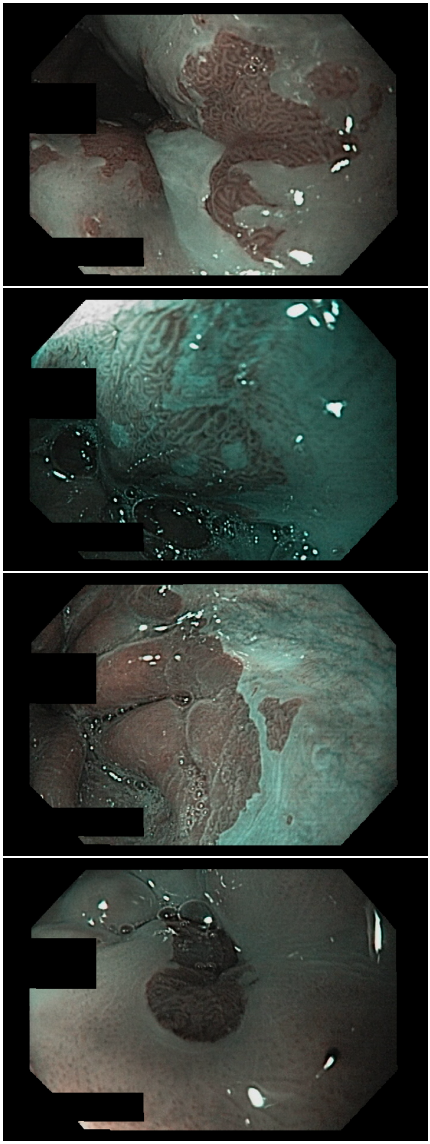

Figure 62: Four images from the dataset, showing pale squamous epithelium and pink metaplasia, masked to remove personal information.

and its manually defined ground truth are shown in Fig. 62 and Fig. 63, respectively.

A.4 Methods

This method aims at the construction of a segmentation mask identifying a candidate region of the image with the highest possibility of being premalignant, according to its texture and color. This is performed by processing the image with a computer vision technique called superpixel segmentation, in particular using the SLIC implementation [82]. The purpose of this process is to create clusters of spatially connected pixels exhibiting similar color and texture. Each cluster is then analyzed, and 8 features are extracted from each of them, to be fed to an ensemble of 50 decision trees, trained using 10-fold cross validation, with a randomly selected 70% of the dataset as training set and the remaining 30% as testing set at each step of the cross-validation procedure.

A.4.1 Superpixel segmentation

As pre-processing steps, each RGB image of the dataset was first edited to remove labels and personal information about each patient and exam by applying a binary mask, as shown in Fig. 62. Then, each image was normalized, transforming each color channel to zero mean and unitary variance. Segmentation via superpixel is then performed by grouping pixels into perceptually meaningful atomic regions, used to replace the rigid structure of the pixel grid. Many computer vision algorithms use superpixels as their building blocks [83; 84], given their straightforwardness and the ease of their implementation. A commonly used superpixel implementation is the Simple Linear Iterative Clustering (SLIC) [82]: this implementation, based on k-means clustering, is fast to compute, memory efficient, simple to use, and outputs superpixels that adhere well to image boundaries. SLIC implementation clusters pixels of the image in the combined five-dimensional color and image plane space to efficiently generate compact and nearly uniform superpixels, imposing a degree

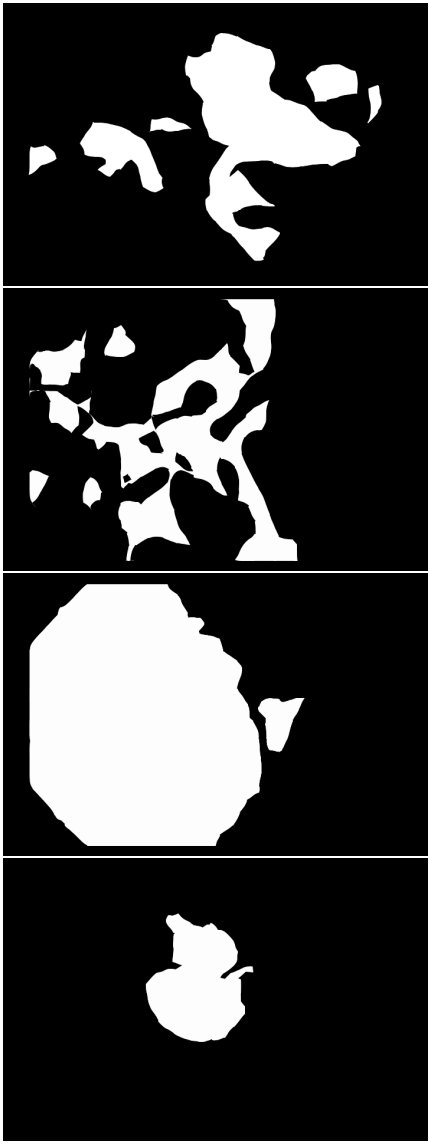

Figure 63: Manually defined ground truth for each of the four images of the dataset showed in the previous Figure.

of spatial regularization to extracted regions. Briefly, the method starts with an initialization step, where all cluster centers are sampled on a regular grid. Centers are then moved to the lowest grayscale gradient position in a 3×3 neighborhood, to avoid centering a superpixel on an edge or on a noisy pixel. Then, a search region is defined for each center, and each pixel belonging to each region is assigned to the closest center. After this, an update step adjusts the clusters' centers to be the mean five-dimensional vector of all pixels belonging to each cluster. A post-processing steps assigns disjoint pixels to nearby superpixels to enforce connectivity. Two different images from the dataset, with superpixels superimposed, are shown in Fig. 64. This step has been implemented with MATLAB R2015b, using an implementation of SLIC superpixels by *vlfeat* [85]. This technique only requires two parameters to set: the desired size of each superpixel N and a regularization parameter λ, that tweaks the smoothness of their contours.

A.4.2 Feature extraction

To be able to distinguish among normal and metaplastic superpixels, some features needs to be defined on the basis of the meaning and appearance of normal and metaplastic regions. A total of 8 groups of features have been identified from each superpixel as such:

- Mean intensities I_r, I_g and I_b: color is clearly among the peculiar differences among the distinction from normal to metaplastic epithelium. Hence, the three different average intensities of the image in RGB color space were selected as features;

- Mean intensity after Top-Hat filtering, I_{th}: Morphological top-hat filtering is used to find the brightest spots in images with not uniform background, correcting the effects of an uneven illumination. Top-hat filtering results in images with smoother and darker image intensity, but with an enhancement of biological patterns of interest as vascularization and tissue rugosity. Since this filtering can only be performed to monochrome images, this feature was extracted

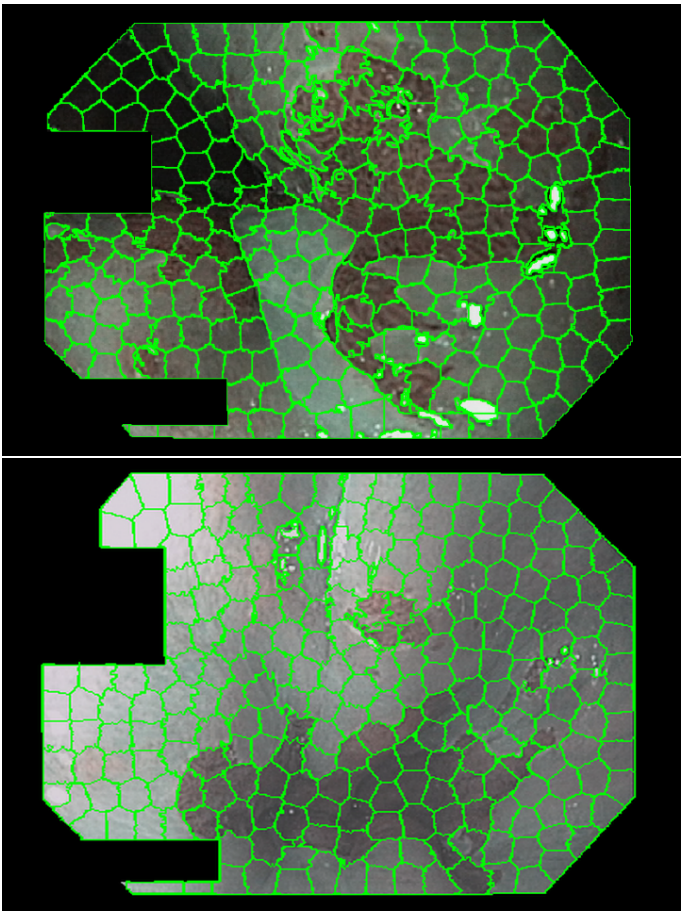

Figure 64: Two images from the dataset, with superpixel segmentation superimposed. As this figure shows, superpixels cluster together pixels exhibiting similar texture and are spatially close one another.

from the red channel, the most discriminative of the three among normal and metaplastic tissue;

- Mean intensity after Entropy Filtering, I_{en}: Entropy filtering calculates, for each pixel, the entropy value of a neighborhood around that pixel. Entropy is a measure of randomness, indicating how much information (i.e., contrast, relative to greyscale intensity variations) is encoded in an image.

- Mean intensity after Range Filtering, I_{ra}: Range filtering enhances regions exhibiting sudden intensity changes among the image. Given this, this filtering emphasizes patterns and borders of the object depicted in the images. Range and entropy filtering are useful, since the two extracted features are statistical measures of the texture of an image, providing information about the local variability of their pixels.

- Contrast C_S and Homogeneity H_S from the Gray Level Co-Occurrence Matrix (GLCM): GLCM is a statistical method of examining texture considering the spatial relationship of pixels. It calculates how often pairs of pixels with specified values and spatial locations occur in an image, building an 8×8 occurrence matrix. Extracting statistical measures from this matrix provides information about the specific texture. From this analysis, contrast (local variations in the GLCM) and homogeneity (how close the distribution of the elements in the GLCM is to its diagonal values) measures have been included in the feature set.

A.4.3 Classification with random forests

For each superpixel, the probability of it being part of a villus fold is computed as the score of a binary random forest [101] classifier using 50 classification trees. To obtain a robust estimation of the performances of the proposed classification method, a cross-validation approach has been chosen. For each step of the 10 folds, the random forest has been trained

on a random sample consisting of 70% of the images in the dataset, and tested on the remaining 30%.

A.5 Results

Using 116 RGB images, the method achieves 83.9% accuracy, with 79.2% sensitivity and 87.3% specificity. In Fig. 65 a comparison among manually obtained ground truth and automatic results is shown for one of the images in the dataset. As shown, the detected metaplastic region strongly correlates with the labeled metaplastic mucosa. False positives are included in the metaplastic mucosa due to light reflectance and/or shadowing artifacts, which are more similar in texture to metaplastic tissue than to squamous epithelium.

A.6 Conclusions

In this work a fast, efficient, stable and reliable method for the automatic detection of metaplastic regions in NBI endoscopic images has been presented, using SLIC superpixel segmentation and random forest classification. This tool will be tested on different dataset of images, from different fields of medical imaging having the same needs as NBI. With more quantitative tools, experts will have a chance to perform targeted biopsies instead of using the four-quadrant standard protocol, improving the accuracy of the exam and improving the examination's accuracies.

Acknowledgements

Thanks to Istituto Oncologico Veneto (http://www.ioveneto.it) for providing the dataset used in this work.

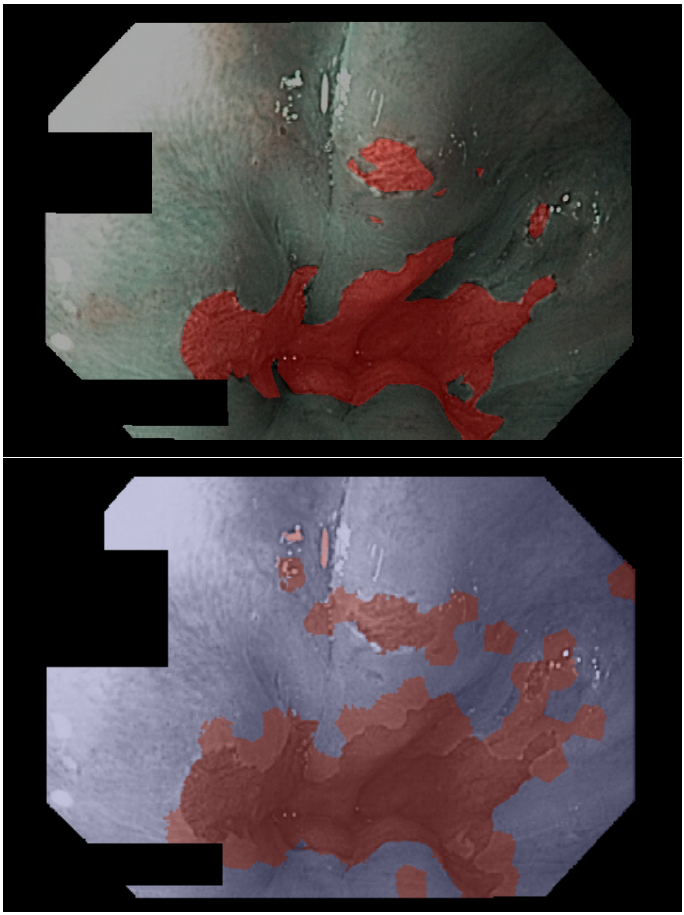

Figure 65: Comparison among metaplastic region according to ground truth (top) and the proposed algorithm (bottom).

Appendix B

Classification of gastric chromoendoscopy images

B.1 Summary

Barrett's esophagus (BE) is a precancerous complication of gastroesophageal reflux disease in which normal stratified squamous epithelium lining the esophagus is replaced by intestinal metaplastic columnar epithelium. Repeated endoscopies and multiple biopsies are often necessary to establish the presence of intestinal metaplasia. Narrow Band Imaging (NBI) is an imaging technique commonly used with endoscopies that enhances the contrast of vascular pattern on the mucosa. A computer-based method for the automatic normal/metaplastic classification of endoscopic NBI images is presented. Superpixel segmentation is used to identify and cluster pixels belonging to uniform regions. From each uniform clustered region of pixels, eight features maximizing differences among normal and metaplastic epithelium are extracted for the classification step. For each superpixel, the three mean intensities of each color channel are firstly selected as features. Three added features are the mean intensities for each superpixel after separately applying to the red-channel image three different morphological filters (top-hat filtering, entropy filtering and range filtering). The last two features require the computation of

the Grey-Level Co-Occurrence Matrix (GLCM), and are reflective of the contrast and the homogeneity of each superpixel. The classification step is performed using an ensemble of 50 classification trees, with a 10-fold cross-validation scheme by training the classifier at each step on a random 70% of the images and testing on the remaining 30% of the dataset. Sensitivity and Specificity are respectively of 79.2% and 87.3%, with an overall accuracy of 83.9%.

B.2 Introduction

Chromoendoscopy (CH) is a gastroenterology imaging modality that involves the staining of tissues with methylene blue, which reacts with the internal walls of the gastrointestinal tract, improving the visual contrast in mucosal surfaces and thus enhancing a doctor's ability to screen precancerous lesions or early cancer. This technique helps identify areas that can be targeted for biopsy or treatment and the focus of this work will be gastric cancer detection. Gastric chromoendoscopy for cancer detection is a highly mature medical field with solid clinical taxonomies, including the most relevant one introduced by Dinis-Ribeiro [102], which is used in this work. According to this taxonomy, CH images are classified into their respective classes based on color, shape and regularity of pit patterns (Figure 1).

Clinical studies show that Group I images are considered normal, Group II cases are considered metaplasia lesions and could lead to cancer lesions. Group III are considered dysplasia lesions. For the purposes of this work, Group I images are considered *normal*, and images from Groups II and III are considered *abnormal*.

The specific goal is to classify the manually segmented region of each image as either *normal* or *abnormal*. The training and test sets have the same size, and each of them is composed by 28 images for the normal class, and 60 images for the abnormal class. Images from the dataset and respective manually defined ground truth are shown in Fig. 67 and Fig. 66, respectively.

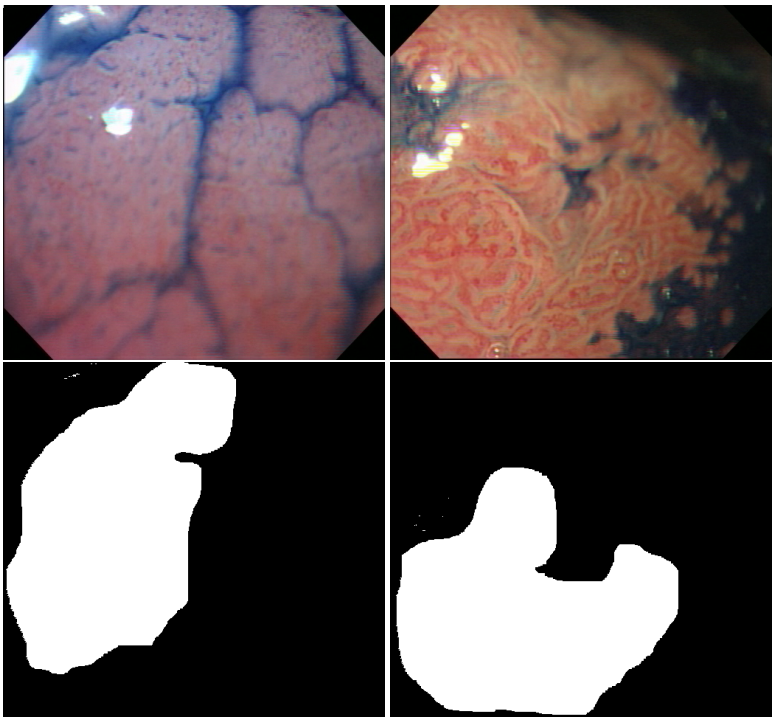

Figure 66: Two images labeled as normal, and their respective manual annotations.

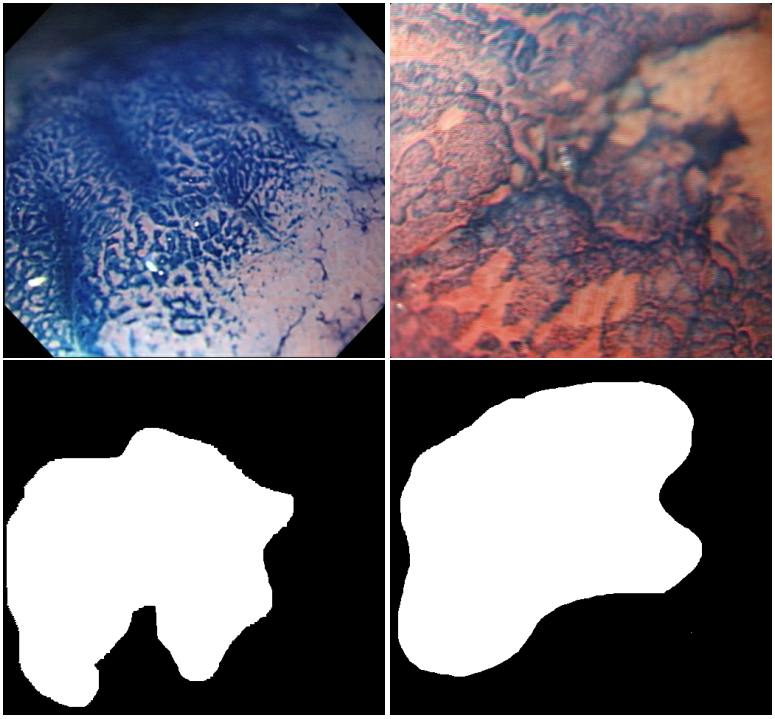

Figure 67: Two images from the dataset labeled as abnormal, and their respective manual annotations.

B.3 Materials

Group I images are considered normal, Group II cases are considered metaplasia lesions and could lead to cancer lesions. The images were acquired using an Olympus GIF-H180 endoscope at the Portuguese Institute of Oncology (IPO) in Porto, Portugal during routine clinical work. Optical characteristics of this endoscope include 140o field of view and four-way angulation (210o up, 90o down and 100o right/left). Annotations were performed independently by two medical experts leading to a gold-standard final annotation which is used here [103].

B.4 Methods

The first step in the proposed method is performed by processing the image with a computer vision technique called superpixel segmentation, using the SLIC implementation [82]. The purpose of this process is to create clusters of spatially connected pixels exhibiting similar texture. Each of the superpixels is then analyzed, and 111 features are extracted from each of them, to be fed to a classifier. The classification step is performed with an ensemble of 50 decision trees. The classifier has been trained on 88 images, and then tested on other different 88 images (for both cases, split in 26 normal and 60 abnormal, with a ratio of approximately 70%-30%). Each image is labeled Normal or Abnormal based on a majority vote among the predicted label of all its superpixels.

B.4.1 Superpixel segmentation

As a pre-processing step for each image, all greyscale values were normalized between 0 and 256, and a median filter was then applied to reduce noise. Segmentation via superpixel is then performed by grouping pixels into perceptually meaningful atomic regions, used to replace the rigid structure of the pixel grid. Many computer vision algorithms use superpixels as their building blocks, given their straightforwardness and the ease of their implementation. A commonly used superpixel im-

plementation is the Simple Linear Iterative Clustering (SLIC) [82]: this implementation, based on k-means clustering, is fast to compute, memory efficient, simple to use, and outputs superpixels that adhere well to image boundaries. SLIC implementation clusters pixels of the image to efficiently generate compact and nearly uniform superpixels, imposing a degree of spatial regularization to extracted regions. This step has been implemented with MATLAB R2015b, using an implementation of SLIC superpixels by *vlfeat* [85]. This technique only requires two parameters to set: the desired size of each superpixel N and a regularization parameter λ, that tweaks the smoothness of their contours. Each region of this image (corresponding to each computed superpixel) is then analyzed for feature extraction.

B.4.2 Feature extraction

A total of 111 features are extracted from the analysis of each superpixel S, 37 for each image color plane:

- Mean intensity μ_S and standard deviation σ_S: greyscale intensity variations are the most basic difference among normal and abnormal tissue;

- Contrast C_S, Energy E_S and Homogeneity H_S from the Gray Level Co-Occurrence Matrix (GLCM): GLCM is a statistical method of examining texture considering the spatial relationship of pixels. It calculates how often pairs of pixels with specified values and spatial locations occur in an image, building a 8×8 occurrence matrix. Extracting statistical measures from this matrix provide information about the specific texture. From this analysis, contrast (local variations in the GLCM), energy (sum of squared elements in GLCM) and homogeneity (how close the distribution of the elements in the GLCM is to its diagonal values) measures have been included in the feature set;

- Histogram of Local Binary Patterns [86] with 32 bins, $hLBP_S$. Local Binary Patterns (LBP) are one of the most descriptive features in

the field of texture classification, and are commonly used in computer vision. They permit the creation of features able to identify different textures in an image. In this work, for each pixel of the image, an 8-bit word is created by comparing its greyscale intensity value with the ones in its 8-neighborhood. Iteratively, starting from a fixed direction, if the central pixel has a grayscale value greater than its neighbor a 1 is encoded in the 8-bit word, a 0 otherwise. When a word has been assigned to each pixel, each word is translated to decimal (0-256). A histogram (32 bins) is then computed for the LBP of pixels in each superpixel, expressing in such way the spectrum of the texture of the selected portion of the image. This is finally added to the feature vector.

B.4.3 Classification with random forests

For each superpixel, the probability of it being normal or abnormal is computed as the score of a binary random forest classifier using 50 classification trees. Accuracy, sensitivity and specificity of the classification step have been computed, in image space.

B.5 Results and discussion

Superpixel parameters were set as $N = 90, \lambda = 0.05$ to obtain a large training dataset and reasonable-sized superpixels, each of them resulting well adherent to image borders. The proposed method reached an average general accuracy of 92.05% in the image space, 86.62% in the superpixel space, respectively over 88 images and 1173 superpixels. A chart showing how accuracy depends on superpixel size is shown in Fig. 69. The smaller the superpixel, the longer the computation time in the feature extraction and training process. A superpixel size of 90 has been chosen, leading reasonable results with a relatively fast training process. As an example, Fig. 68 shows the computational times for feature extraction with some superpixel sizes (on an Intel i7-3630QM @ 2.40 GHz with a Samsung 840 EVO SSD drive and 8 GB RAM).

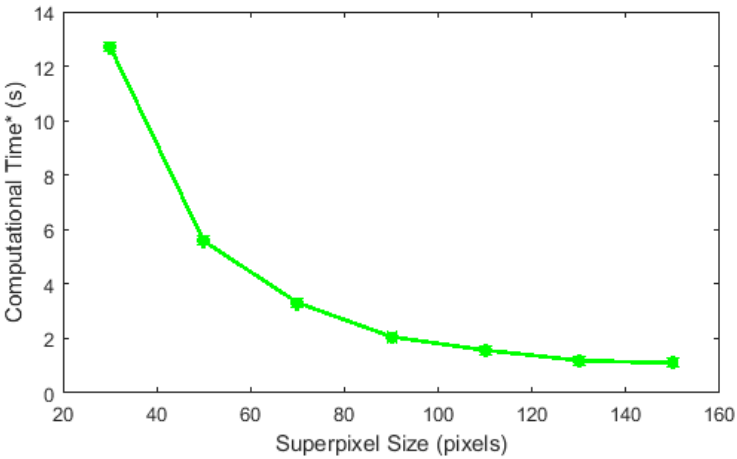

Figure 68: Average computation time for loading the two images (image and ROI) and feature extraction, with varying superpixel size. Times in second per image, obtained by averaging on the full computation on 88 images.

For the chosen superpixel size, detection accuracy on the normal class is 85.71%, while detection accuracy on the abnormal class is 95%. These two metrics in the superpixel domain are, respectively, 79.95% and 90.39%.

A variable that could be tweaked with thought is the ratio among normal and abnormal superpixels needed to classify one image into normal or abnormal. In Fig. 71 it is shown that by requiring less than half superpixels to classify an image as abnormal, general accuracy changes, along with normal and abnormal class accuracies. It is clear from the image that altering this ratio favours the accuracy in one of the two classes (the favoured one), but does not help in general accuracy. The best value is therefore 0.5, as logic suggests. The only drawback of this method is if ROIs are small enough that no superpixels fit inside of them: in such cases, no superpixels from that image are used either for training or testing, leading to a failure in that particular image. In these dataset, this has happened in one case only: abnormal class image number 045 from the training set, shown in Fig. 70. Reducing the superpixels size solved the

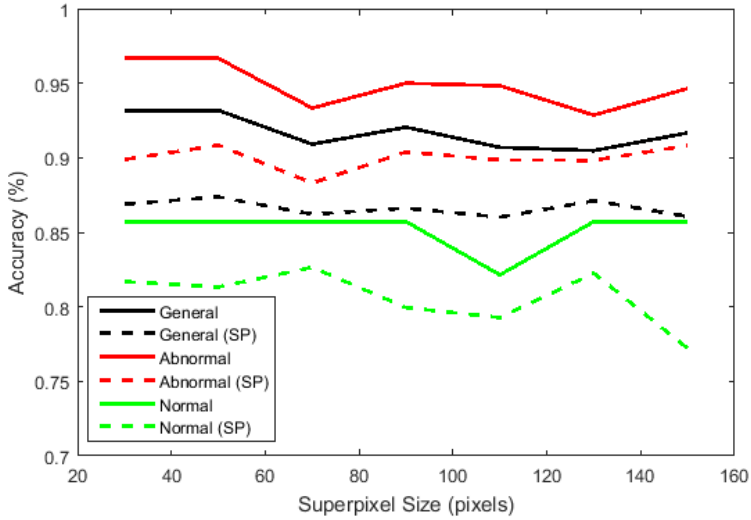

Figure 69: How superpixel size affect average accuracy.

problem. For all computations with superpixel size greater than 70, this particular image gave no contribution in the training process.

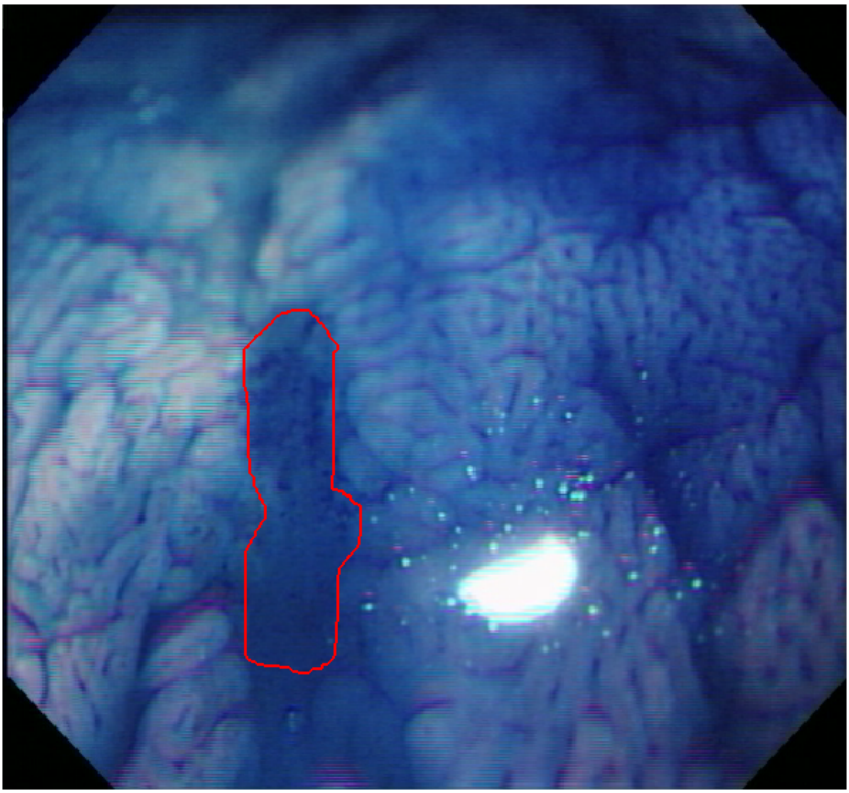

Figure 70: Image 05 and the manual mask superimposed.

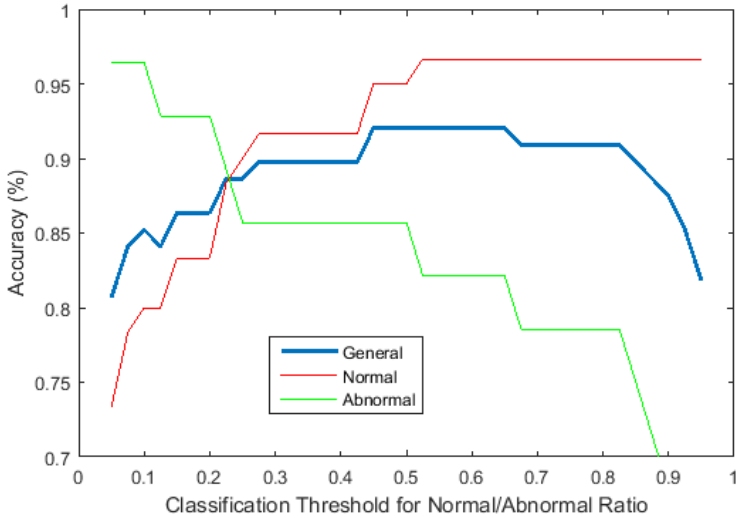

Figure 71: How the ratio between normal/abnormal superpixel used in the final image classification affect average accuracy, once superpixel size has been fixed.

References

[1] Anna M Buchner, Muhammad W Shahid, Michael G Heckman, Murli Krishna, Marwan Ghabril, Muhammad Hasan, Julia E Crook, Victoria Gomez, Massimo Raimondo, Timothy Woodward, Herbert C Wolfsen, and Michael B Wallace. Comparison of probe-based confocal laser endomicroscopy with virtual chromoendoscopy for classification of colon polyps. *Gastroenterology*, 138(3):834–42, mar 2010. 2

[2] V Gómez, A M Buchner, E Dekker, F J van den Broek, A Meining, M W Shahid, M S Ghabril, P Fockens, M G Heckman, and M B Wallace. Interobserver agreement and accuracy among international experts with probe-based confocal laser endomicroscopy in predicting colorectal neoplasia. *Endoscopy*, 42(4):286–91, apr 2010. 2

[3] Michael B Wallace and Ralf Kiesslich. Advances in endoscopic imaging of colorectal neoplasia. *Gastroenterology*, 138(6):2140–50, jun 2010. 2

[4] Sebastian Hegenbart, Andreas Uhl, Andreas Vécsei, and G. Wimmer. Scale invariant texture descriptors for classifying celiac disease. *Medical Image Analysis*, 17(4):458–474, 2013. 3

[5] S. Hegenbart, A. Uhl, and A. Vcsei. On the implicit handling of varying distances and gastrointestinal regions in endoscopic video sequences with indication for celiac disease. In *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, pages 1–6, June 2012. 3

[6] M. Bonamico, P. Mariani, E. Thanasi, M. Ferri, R. Nenna, C. Tiberti, B. Mora, M. C. Mazzilli, and F. M. Magliocca. Patchy villous atrophy of the duodenum in childhood celiac disease. *J. Pediatr. Gastroenterol. Nutr.*, 38(2):204–207, Feb 2004. 3

[7] A. D. Hopper, S. S. Cross, and D. S. Sanders. Patchy villous atrophy in adult patients with suspected gluten-sensitive enteropathy: is a multiple duodenal biopsy strategy appropriate? *Endoscopy*, 40(3):219–224, Mar 2008. 3

[8] S. Hegenbart, A. Uhl, and A. Vcsei. Impact of endoscopic image degradations on lbp based features using one-class svm for classification of celiac disease. In *Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on*, pages 715–720, Sept 2011. 3

[9] Sebastian Hegenbart, Andreas Uhl, and Andreas Vécsei. Survey on computer aided decision support for diagnosis of celiac disease. *Computers in Biology and Medicine*, 65:348–358, 2015. 3, 12

[10] F. Biagi and G. R. Corazza. Mortality in celiac disease. *Nat Rev Gastroenterol Hepatol*, 7(3):158–162, Mar 2010. 10

[11] J. D. Godfrey, T. L. Brantner, W. Brinjikji, K. N. Christensen, D. L. Brogan, C. T. Van Dyke, B. D. Lahr, J. J. Larson, A. Rubio-Tapia, L. J. Melton, A. R. Zinsmeister, R. A. Kyle, and J. A. Murray. Morbidity and mortality among older individuals with undiagnosed celiac disease. *Gastroenterology*, 139(3):763–769, Sep 2010. 10

[12] A. Rubio-Tapia, C. T. Van Dyke, B. D. Lahr, A. R. Zinsmeister, M. El-Youssef, S. B. Moore, M. Bowman, L. J. Burgart, L. J. Melton, and J. A. Murray. Predictors of family risk for celiac disease: a population-based study. *Clin. Gastroenterol. Hepatol.*, 6(9):983–987, Sep 2008. 10

[13] A. Carlsson, D. Agardh, S. Borulf, E. Grodzinsky, I. Axelsson, and S. A. Ivarsson. Prevalence of celiac disease: before and after a national change in feeding recommendations. *Scand. J. Gastroenterol.*, 41(5):553–558, May 2006. 10

[14] L. Castano, E. Blarduni, L. Ortiz, J. Nunez, J. R. Bilbao, I. Rica, P. Martul, and J. C. Vitoria. Prospective population screening for celiac disease: high prevalence in the first 3 years of life. *J. Pediatr. Gastroenterol. Nutr.*, 39(1):80–84, Jul 2004. 10

[15] C. Dube, A. Rostom, R. Sy, A. Cranney, N. Saloojee, C. Garritty, M. Sampson, L. Zhang, F. Yazdi, V. Mamaladze, I. Pan, J. Macneil, D. Mack, D. Patel, and D. Moher. The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology*, 128(4 Suppl 1):57–67, Apr 2005. 10

[16] J. Henker, A. Losel, K. Conrad, T. Hirsch, and W. Leupold. [Prevalence of asymptommatic coeliac disease in children and adults in the Dresden region of Germany]. *Dtsch. Med. Wochenschr.*, 127(28-29):1511–1515, Jul 2002. 10

[17] A. Rubio-Tapia, J. F. Ludvigsson, T. L. Brantner, J. A. Murray, and J. E. Everhart. The prevalence of celiac disease in the United States. *Am. J. Gastroenterol.*, 107(10):1538–1544, Oct 2012. 10

[18] C. Catassi, I. M. Ratsch, E. Fabiani, M. Rossini, F. Bordicchia, F. Candela, G. V. Coppa, and P. L. Giorgi. Coeliac disease in the year 2000: exploring the iceberg. *Lancet*, 343(8891):200–203, Jan 1994. 10

[19] J. R. Brow, F. Parker, W. M. Weinstein, and C. E. Rubin. The small intestinal mucosa in dermatitis herpetiformis. I. Severity and distribution of the small intestinal lesion and associated malabsorption. *Gastroenterology*, 60(3):355–361, Mar 1971. 10

[20] M. Hadjivassiliou, D. S. Sanders, N. Woodroofe, C. Williamson, and R. A. Grunewald. Gluten ataxia. *Cerebellum*, 7(3):494–498, 2008. 10

[21] A. G. Kalayci, Y. Kanber, A. Birinci, L. Yildiz, and D. Albayrak. The prevalence of coeliac disease as detected by screening in children with iron deficiency anaemia. *Acta Paediatr.*, 94(6):678–681, Jun 2005. 10

[22] C. Mulder, S. van Weyenberg, and M. Jacobs. Celiac disease is not yet mainstream in endoscopy. *Endoscopy*, 42(3):218–9, 2010. 11

[23] A. Tursi, W. Elisei, G. M. Giorgetti, G. Brandimarte, and F. Aiello. Complications in celiac disease under gluten-free diet. *Dig. Dis. Sci.*, 54(10):2175–2182, Oct 2009. 11

[24] A. Fasano and C. Catassi. Current approaches to diagnosis and treatment of celiac disease: an evolving spectrum. *Gastroenterology,*, 120(3):636–51, 2001. 11

[25] EJ. Ciaccio, SK. Lewis, and PH. Green. Detection of villous atrophy using endoscopic images for the diagnosis of celiac disease. *Digestive Diseases and Sciences*, 58(8):1167–9, 2013. 11

[26] M. Bonamico, P. Mariani, E. Thanasi, M. Ferri, R. Nenna, C. Tiberti, B. Mora, MC. Mazzilli, and FM. Magliocca. Patchy villous atrophy of the duodenum in childhood celiac disease. *J Pediatr Gastroenterol Nutr*, 38(2):204–7, 2004. 11

[27] KEA. Lundin, EM. Nilsen, HG. Scott, EM. Løberg, A. Gjøen, J. Bratlie, V. Skar, E. Mendez, A. Løvik, and K. Kett. Oats induced villous atrophy in coeliac disease. *Gut*, 52(11):1649–52, 2003. 11

[28] S. Hüe, JJ. Mention, RC. Monteiro, S. Zhang, C. Cellier, J. Schmitz, V. Verkarre, N. Fodil, S. Bahram, N. Cerf-Bensussan, and S. Caillat-Zucman. A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease. *Immunity*, 21(3):367–77, 2004. 11

[29] EJ. Ciaccio, G. Bhagat, SK. Lewis, and PH. Green. Quantitative image analysis of celiac disease. *World J Gastroenterol*, 21(9):2577–81, 2015. 11

[30] G. Oberhuber, G. Granditsch, and H. Vogelsang. The histopathology of coeliac disease: time for a standardized report scheme for pathologists. *Eur J Gastroenterol Hepatol*, 11(10):1185–1194, Oct 1999. 12

[31] M. N. Marsh. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology*, 102(1):330–354, Jan 1992. 12

[32] A. Mubarak, P. Nikkels, R. Houwen, and F. Ten Kate. Reproducibility of the histological diagnosis of celiac disease. *Scand. J. Gastroenterol.*, 46(9):1065–1073, Sep 2011. 12

[33] C. Arguelles-Grande, C. A. Tennyson, S. K. Lewis, P. H. Green, and G. Bhagat. Variability in small bowel histopathology reporting between different pathology practice settings: impact on the diagnosis of coeliac disease. *J. Clin. Pathol.*, 65(3):242–247, Mar 2012. 12

[34] J. Taavela, O. Koskinen, H. Huhtala, M. L. Lahdeaho, A. Popp, K. Laurila, P. Collin, K. Kaukinen, K. Kurppa, and M. Maki. Validation of morphometric analyses of small-intestinal biopsy readouts in celiac disease. *PLoS ONE*, 8(10):e76163, 2013. 12

[35] G. R. Corazza, V. Villanacci, C. Zambelli, M. Milione, O. Luinetti, C. Vindigni, C. Chioda, L. Albarello, D. Bartolini, and F. Donato. Comparison of the interobserver reproducibility with different histologic criteria used in celiac disease. *Clin. Gastroenterol. Hepatol.*, 5(7):838–843, Jul 2007. 12

[36] G. Cammarota, G. A. Pirozzi, A. Martino, G. Zuccala, R. Cianci, L. Cuoco, V. Ojetti, M. Landriscina, M. Montalto, F. M. Vecchio, G. Gasbarrini, and A. Gasbarrini. Reliability of the "immersion technique" during routine upper endoscopy for detection of abnormalities of duodenal villi in patients with dyspepsia. *Gastrointest. Endosc.*, 60(2):223–228, Aug 2004. 12

[37] G. Cammarota, P. Cesaro, A. Martino, G. Zuccala, R. Cianci, E. Nista, L. M. Larocca, F. M. Vecchio, A. Gasbarrini, and G. Gasbarrini. High accuracy and cost-effectiveness of a biopsy-avoiding endoscopic approach in diagnosing coeliac disease. *Aliment. Pharmacol. Ther.*, 23(1):61–69, Jan 2006. 12

[38] MC. Arrieta, L. Bistritz, and JB. Meddings. Alterations in intestinal permeability. *Gut*, 55(10):1512–1520, Oct 2006. 12

[39] Visser, J. et al. Tight junctions, intestinal permeability, and autoimmunity: celiac disease and type 1 diabetes paradigms. *Ann. N. Y. Acad. Sci.*, 1165:195–205, May 2009. 12

[40] C. W. Teshima, L. A. Dieleman, and J. B. Meddings. Abnormal intestinal permeability in Crohn's disease pathogenesis. *Ann. N. Y. Acad. Sci.*, 1258:159–165, Jul 2012. 12

[41] C. Buning, N. Geissler, M. Prager, A. Sturm, D. C. Baumgart, J. Buttner, S. Buhner, V. Haas, and H. Lochs. Increased small intestinal permeability in ulcerative colitis: rather genetic than environmental and a risk factor for extensive disease? *Inflamm. Bowel Dis.*, 18(10):1932–1939, Oct 2012. 12

[42] Y. Ko, R. Butcher, and R. W. Leong. Epidemiological studies of migration and environmental risk factors in the inflammatory bowel diseases. *World J. Gastroenterol.*, 20(5):1238–1247, Feb 2014. 13

[43] J. K. Hou, B. Abraham, and H. El-Serag. Dietary intake and risk of developing inflammatory bowel disease: a systematic review of the literature. *Am. J. Gastroenterol.*, 106(4):563–573, Apr 2011. 13

[44] N. A. Molodecky, I. S. Soon, D. M. Rabi, W. A. Ghali, M. Ferris, G. Chernoff, E. I. Benchimol, R. Panaccione, S. Ghosh, H. W. Barkema, and G. G. Kaplan. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*, 142(1):46–54, Jan 2012. 13

[45] A. N. Ananthakrishnan, L. M. Higuchi, E. S. Huang, H. Khalili, J. M. Richter, C. S. Fuchs, and A. T. Chan. Aspirin, nonsteroidal anti-inflammatory drug use, and risk for Crohn disease and ulcerative colitis: a cohort study. *Ann. Intern. Med.*, 156(5):350–359, Mar 2012. 13

[46] A. C. von Roon, G. Reese, J. Teare, V. Constantinides, A. W. Darzi, and P. P. Tekkis. The risk of cancer in patients with Crohn's disease. *Dis. Colon Rectum*, 50(6):839–855, Jun 2007. 13

[47] H. J. Freeman. Colorectal cancer risk in Crohn's disease. *World J. Gastroenterol.*, 14(12):1810–1811, Mar 2008. 13

[48] Shanahan F. Ulcerative colitis. In Hawkey CJ, Bosch J, Richter JE, Garcia-Tsao G, and Chan FKL, editors, *Textbook of Clinical Gastroenterology and Hepatology, 2nd edition*, chapter 49. Wiley, 2012. 14

[49] A. Khalaf, C. L. Hoad, R. C. Spiller, P. A. Gowland, G. W. Moran, and L. Marciani. Magnetic resonance imaging biomarkers of gastrointestinal motor function and fluid distribution. *World J Gastrointest Pathophysiol*, 6(4):140–149, Nov 2015. 15

[50] D. Martin, R. Danrad, K. Herrmann, R. Semelka, and S. Hussain. Magnetic Resonance Imaging of the Gastrointestinal Tract. *Topics in Magnetic Resonance Imaging*, 16(1):77–98, Feb 2005. 16

[51] H. Brambs and M.S. Juchems. Virtual endoscopy using CT scan. *Minimally Invasive Therapy and Allied Technologies*, 12(5):207–216, 2003. 16

[52] Autofluorescence Imaging from ASGE. `http://www.asge.org/uploadedFiles/Publications_and_Products/Technology_Reviews/Autofluorescence%20imaging.pdf`, 2016. 21, 22

[53] G. W. Falk. Autofluorescence endoscopy. *Gastrointest. Endosc. Clin. N. Am.*, 19(2):209–220, Apr 2009. 21

[54] American Society for Gastrointestinal Endoscopy. `http://www.asge.org/press/press.aspx?id=8140`, 2016. 22

[55] M. Bhattarai, P. Bansal, and Y. Khan. Longest duration of retention of video capsule: A case report and literature review. *World J Gastrointest Endosc*, 5(7):352–355, Jul 2013. 23

[56] R. Eliakim. Wireless capsule video endoscopy: three years of experience. *World J. Gastroenterol.*, 10(9):1238–1239, May 2004. 23

[57] P. E. Paull, B. J. Hyatt, W. Wassef, and A. H. Fischer. Confocal laser endomicroscopy: a primer for pathologists. *Arch. Pathol. Lab. Med.*, 135(10):1343–1348, Oct 2011. 25

[58] Pentax. `http://www.pentaxmedical.com`, 2016. 25

[59] Mauna Kea Technologies. `http://www.maunakeatech.com`, 2016. 25

[60] M. B. Wallace and P. Fockens. Probe-based confocal laser endomicroscopy. *Gastroenterology*, 136(5):1509–1513, May 2009. 25, 27

[61] M. C. Pierce, D. J. Javier, and R. Richards-Kortum. Optical contrast agents and imaging systems for detection and diagnosis of cancer. *Int. J. Cancer*, 123(9):1979–1990, Nov 2008. 25

[62] S. V. Kantsevoy, D. G. Adler, J. D. Conway, D. L. Diehl, F. A. Farraye, V. Kaul, S. R. Kethu, R. S. Kwon, P. Mamula, S. A. Rodriguez, and W. M. Tierney. Confocal laser endomicroscopy. *Gastrointest. Endosc.*, 70(2):197–200, Aug 2009. 27

[63] H. Neumann, R. Kiesslich, M. B. Wallace, and M. F. Neurath. Confocal laser endomicroscopy: technical advances and clinical applications. *Gastroenterology*, 139(2):388–392, Aug 2010. 27

[64] A. L. Polglase, W. J. McLaren, S. A. Skinner, R. Kiesslich, M. F. Neurath, and P. M. Delaney. A fluorescence confocal endomicroscope for in vivo microscopy of the upper- and the lower-GI tract. *Gastrointest. Endosc.*, 62(5):686–695, Nov 2005. 27

[65] R. Kiesslich, J. Burg, M. Vieth, J. Gnaendiger, M. Enders, P. Delaney, A. Polglase, W. McLaren, D. Janell, S. Thomas, B. Nafe, P. R. Galle, and M. F. Neurath. Confocal laser endoscopy for diagnosing intraepithelial neoplasias and colorectal cancer in vivo. *Gastroenterology*, 127(3):706–713, Sep 2004. 27

[66] M. Wainwright. Acridine-a neglected antibacterial chromophore. *J. Antimicrob. Chemother.*, 47(1):1–13, Jan 2001. 27

[67] R. Kiesslich, M. Goetz, E. M. Angus, Q. Hu, Y. Guan, C. Potten, T. Allen, M. F. Neurath, N. F. Shroyer, M. H. Montrose, and A. J. Watson. Identification of epithelial gaps in human small and large intestine by confocal endomicroscopy. *Gastroenterology*, 133(6):1769–1778, Dec 2007. 27, 70

[68] S. Foersch, R. Kiesslich, M. J. Waldner, P. Delaney, P. R. Galle, M. F. Neurath, and M. Goetz. Molecular imaging of VEGF in gastrointestinal cancer in vivo using confocal laser endomicroscopy. *Gut*, 59(8):1046–1055, Aug 2010. 27

[69] Liedlgruber M. and Uhl A. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: A review. *IEEE Reviews in Biomedical Imaging**********, 4:73–88, 2011. 28, 30

[70] S. Hwang, J. Oh, W. Tavanapong, J. Wong, and P. C. de Groen. Polyp detection in colonoscopy video using elliptical shape feature. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 2, pages II – 465–II – 468, Sept 2007. 28

[71] G. Oberhuber. Histopathology of celiac disease. *Biomed. Pharmacother.*, 54(7):368–372, Aug 2000. 34

[72] B. C. Dickson, C. J. Streutker, and R. Chetty. Coeliac disease: an update for pathologists. *J. Clin. Pathol.*, 59(10):1008–1016, Oct 2006. 34

[73] J. Chang, M. Ip, M. Yang, B. Wong, T. Power, L. Lin, W. Xuan, TG. Phan, and R. Leong. The learning curve, interobserver, and intraobserver agreement of endoscopic confocal laser endomicroscopy in the assessment of mucosal barrier defects. *Gastrointest. Endosc.*, Sep 2015. 34, 45

[74] IS. Brown, DC. Whiteman, and GY. Lauwers. Foveolar type dysplasia in barrett esophagus. *Modern Pathology*, 23:834–43, 2010. 34

[75] R. Kiesslich, CA. Duckworth, D. Moussata, A. Gloeckner, LG. Lim, M. Goetz, DM. Pritchard, PR. Galle, MF. Neurath, and AJ. Watson. Local barrier dysfunction identified by confocal laser endomicroscopy predicts relapse in inflammatory bowel disease. *Gut*, 61(8):1146–1153, Aug 2012. 34, 70

[76] N. Ali. Understanding celiac disease: An introduction for patients and caregivers. *Rowman and Littlefield Publishers*, 2014. 34

[77] H. Mirzaei, C. Fung, J. Chang, and R. Leong. Confocal laser endomicroscopy is equivalent to histology in the detection of gluten-free diet efficacy in coeliac disease. *Journal of Gastroenterology and Hepatology, Special Issue: Australian Gastroenterology Week 2014 Gold Coast Convention and Exhibition Centre, Broadbeach, Queensland, 22-24 October 2014*, 29(S2):53, 2014. 34

[78] R. Leong, N. Nguyen, C. Meredith, S. Al-Sohaily, D. Kukic, P. Delaney, E. Murr, J. Yong, N. Merrett, and A. Biankin. In vivo confocal endomicroscopy in the diagnosis and evaluation of celiac disease. *Gastroenterology*, 135(6):1870–6, 2008. 36

[79] L. Vincent. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Transactions on Image Processing*, 2(2):176–201, 1993. 40

[80] RA. Finkel and J.L. Bentley. Quad trees a data structure for retrieval on composite keys. *Acta Informatica*, 21(XI):1–9, 1974. 40

[81] SJ. Orfanidis. *Introduction to Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ, 1996. 43

[82] Achanta, R. et al. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell*, 34(11):2274–2282, Nov 2012. 48, 50, 115, 127, 128

[83] B. Fulkerson, A. Vedaldi, and S. Soatto. Class segmentation and object localization with superpixel neighborhoods. *Computer Vision, 2009 IEEE 12th International Conference on*, pages 670–677, 2009. 50, 115

[84] Li, Y. et al. Lazy snapping. *ACM Transactions on Graphics (SIGGRAPH)*, 3(23):303–308, 2004. 50, 115

[85] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008. 52, 117, 128

[86] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. *Computer Vision - ECCV 2000: 6th European Conference on Computer Vision Dublin, Ireland, June 26 – July 1, 2000 Proceedings, Part I*, chapter Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns, pages 404–420. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000. 54, 128

[87] JS. Lim. Two-dimensional signal and image processing. *Englewood Cliffs, NJ, Prentice Hall*, page 548, 1990. 60

[88] F. Aurenhammer. Voronoi diagrams: a survey of a fundamental geometric data structure. *ACM Computing Surveys*, 23(3):345–405, 1991. 62

[89] W. J. Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950. 63

[90] Y. Guan, A. J. Watson, A. M. Marchiando, E. Bradford, L. Shen, J. R. Turner, and M. H. Montrose. Redistribution of the tight junction protein ZO-1 during physiological shedding of mouse intestinal epithelial cells. *Am. J. Physiol., Cell Physiol.*, 300(6):C1404–1414, Jun 2011. 70

[91] J. J. Liu, K. Wong, A. L. Thiesen, S. J. Mah, L. A. Dieleman, B. Claggett, J. R. Saltzman, and R. N. Fedorak. Increased epithelial gaps in the small intestines of patients with inflammatory bowel disease: density matters. *Gastrointest. Endosc.*, 73(6):1174–1180, Jun 2011. 70

[92] NM. Barrett. Chronic peptic ulcer of the oesophagus and 'oesophagitis. *B.R. J. Surg.*, pages 175–182, 1950. 112

[93] KB Dunbar and SJ Spechler. Controversies in barrett's esophagus. *Mayo Clinic Proc.*, 89:973–984, 2014. 112

[94] G Falk. Barrett's esophagus. *Gastroenterology*, pages 1569–1591, 2002. 112

[95] J Mannath and K Ragunath. Reflux and barrett's disease. *Endoscopy*, pages 34–37, 2010. 112

[96] P. Sharma. Barrett's esophagus. *N. Engl. J. Med.*, pages 2548–2556, 2009. 112

[97] Ishimura, N. et al. Barrett's esophagus: endoscopic diagnosis. *Ann. N.Y. Acad. Sci.*, pages 53–75, 2011. 112

[98] Baldaque-Silva, F. et al. Endoscopic assessment and grading of barrett's esophagus using magnification endoscopy and narrow band imaging: impact of structured learning and experience on the accuracy of the amsterdam classification system. *Scand J Gastroenterol.*, 48:160–167, Feb 2013. 112

[99] SJ. Spechler. Barrett's esophagus and risk of esophageal cancer: a clinical review. *JAMA*, 310:627–636, 2013. 112

[100] Spechler, SJ. et al. American gastroenterological association technical review on the management of barrett's esophagus. *Gastroenterology*, 310:627–636, 2013. 113

[101] Breiman, L. Random forests. *Machine Learning*, 45:5–32, 2001. 119

[102] M. Dinis-Ribeiro. *Clinical, endoscopic and laboratorial assessment of patients with associated lesions to gastric adenocarcinoma*. PhD thesis, Faculdade de Medicina da Universidade do Porto, 2005. 124

[103] F. Riaz, F. B. Silva, M. D. Ribeiro, and M. T. Coimbra. Invariant Gabor texture descriptors for classification of gastroenterology images. *IEEE Trans Biomed Eng*, 59(10):2893–2904, Oct 2012. 127